MASS–SPECTROMETRY BASED METABOLOMIC APPROACHES TO CHARACTERIZE PLANT PATHOGEN RESISTANCE IN CEREALS AND THE FLAVONOLS MODIFYING DEVELOPMENT IN *ARABIDOPSIS THALIANA*

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Rahel Ruth Bucher

von

Altstätten SG

Promotionskomitee

Prof. Dr. Stefan Bienz (Vorsitz)

PD Dr. Laurent Bigler (Leitung der Dissertation)

Prof. Dr. Beat Keller

Zürich, 2017

Table of Contents

— Appendix —

Summary

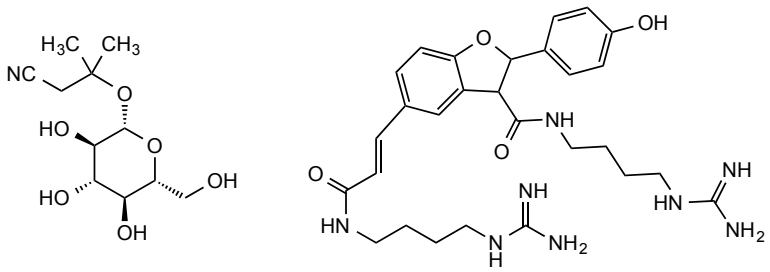
Metabolomics aims at the comprehensive analysis of the sum of all metabolites, and the ultimate goal is to identify and quantify the structurally diverse small molecules in a cell or an organism such as a plant. Metabolomics is an increasingly important tool in plant science that allows to assess plant stress response, to characterize phenotypes, to assist plant breeding and to generally gain a better understanding of cellular systems. The objective of this thesis was to develop dedicated mass-spectrometry based metabolomics methods. A first untargeted approach was designed for the detection of differences in metabolic fingerprints, and was applied to characterize fungal pathogen resistance in cereals. A second targeted approach was designed for monitoring defined metabolites belonging to a distinct metabolic pathway and was utilized for the analysis of a specific metabolite class, flavonols, in the model plant *Arabidopsis thaliana*.

We have successfully developed methods within plant metabolomics workflows for the investigation of the resistance gene *Lr34*, thereby combining efforts of plant biology and analytical chemistry. The metabolic characterization allowed to recognize characteristic differences between control plants and *Lr34* plants, to identify responsible metabolites and to embed the metabolites into pathways for biological interpretation.

Introduction of durable resistance genes in crops is an important strategy to prevent yield loss caused by pathogens and to maintain food security. The *Lr34* resistance gene originating from wheat (*Triticum aestivum*) durably confers resistance to four major fungal biotic pathogens: leaf rust, stripe rust, stem rust and powdery mildew. While *Lr34* in wheat only confers resistance in adult plants, it is functionally transferable to barley (*Hordeum vulgare*) and rice (*Oryza sativa*) where *Lr34* is expressed at seedling stage, thus providing suitable plant material for metabolomics within a shorter time range. The molecular resistance mechanism of *Lr34*, encoding for an adenosine triphosphate-binding cassette transporter, is not known yet. The overall aim of the multi-disciplinary project was to improve the understanding of the molecular function and the defense response of durable disease resistance in cereals. The metabolomic characterization of plants containing the disease resistance gene *Lr34* presented in this thesis was performed in two steps.

In a first series of experiments, metabolic fingerprints of control plants and transgenic *Lr34* barley plants were compared by ultra-high performance liquid chromatography-mass spectrometry (UHPLC-MS), with focus on secondary metabolites. Plants were grown on

hydroponic solution to control the growth conditions as tightly as possible, which was hypothesized to induce less biological variation. Significantly altered metabolic features were selected for metabolite annotation, and two groups of upregulated metabolites had been identified. Based on high-resolution mass spectrometry (HR-MS) and tandem mass spectrometry (MS/MS) data, we were able to annotate hordatines, which are barley-specific hydroxycinnamic acid amines, and cyano glycoside derivatives. The biological function of cyano glycosides in barley is not yet fully understood, but hordatines are known to be involved in defense response and are induced upon fungal pathogen attack. Likely, the metabolites represent indirect responses caused by the substrate of the LR34 transporter, because only a partial upregulation was observed that was assumed to be influenced by the strong leaf-tip necrosis phenotype in barley.



Scheme 1: Classes of cyano glycoside derivatives and hordatines, upregulated metabolites in *Lr34* barley, exemplified with epidermin and hordatine A, respectively.

A second series of experiments focused on the investigation of the metabolic response of *Lr34* containing plants. The aim was to account for robustness of plant growth, to compare effects of *Lr34* in different cereal species and to circumvent the effects of the leaf-tip necrosis phenotype by investigating barley and rice genotypes with low and high expression levels of *Lr34*. Therefore, the design of the investigation was expanded towards different cereal species, different growth conditions involving hydroponic, mock- and fungal pathogen-infected experiments and additional classes of metabolites. Additionally, metabolic profiles of field-grown *Lr34* wheat were used to investigate the resistance gene under natural conditions. A broad range of components including primary and secondary metabolites, lipids and selected plant hormones were analyzed by a combination of UHPLC- and gas chromatography-MS based methods.

The main metabolic differences were found in high-expressing *Lr34* barley. In rice and wheat, the glyoxylate cycle was induced and the according metabolites isocitrate down-

and succinate upregulated. A possible metabolic outcome of the glyoxylate cycle and the subsequent gluconeogenesis is glucose, which might be involved in the metabolic response of *Lr34* as signaling molecule or as nutrient for the biotrophic fungi, which feed on sugars of the living plant. Further, a group of downregulated secondary metabolites was putatively annotated based on HR-MS and MS/MS data as *C*–glycosylated flavones in *Lr34* rice and barley. Flavonoids are known as pathogen defensive compounds and their biosynthesis was enhanced in *Lr34* barley on mRNA level, contrasting to the results from metabolomics.

Overall, this work revealed insights into metabolic pathways involved in the durable multi-pathogen resistance *Lr34* and contributed to gain a deeper understanding of the metabolic response of fungal pathogen resistance in the cereals barley, rice and wheat. Generally, the biological interpretation was complicated by difficulties to distinguish between effects that were directly related to the LR34 transporter and indirect effects caused by the strong leaf-tip necrosis phenotype in barley.

In the second project, we developed an improved UHPLC-MS method for the targeted analysis of flavonols to investigate their biological functions related to plant growth. This included the assignment of previously found and new flavonols.

Flavonoids represent an important group of secondary plant metabolites with multiple of functions (e.g. UV protection, hormone transport, cell development, gene expression). The model plant *Arabidopsis thaliana* contains a flavonoid subgroup termed flavonols is composed of kaempferol, quercetin and isorhamnetin cores that are modified with glucose and rhamnose units. LC-MS is the method of choice to characterize flavonols in complex plant extracts. A previously developed method for targeted analysis of flavonols in *A. thaliana* extracts was improved for better selectivity, sensitivity, mass accuracy, and speed. Flavonol annotations were reassigned in particular based on biological mutants and on calculations of chemical formulas based on HR-MS, and on MS/MS spectra.

Flavonol profiles obtained from the new method were successfully used to study the role of flavonols modifying auxin transport, cell growth and plant development. Thus, the improved method enables an accurate targeted analysis of flavonols in plants, and is a versatile tool to study their complex and diverse functions.

Zusammenfassung

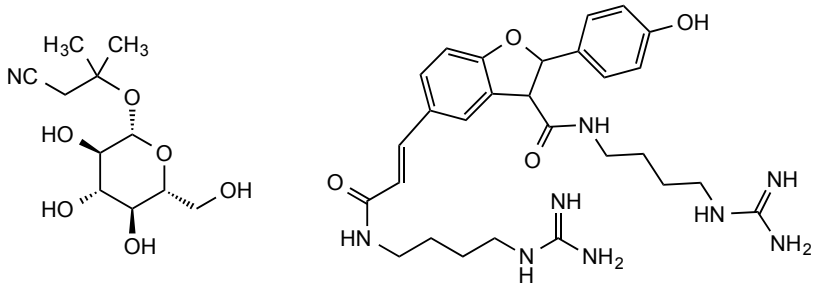
Metabolomics zielt auf die umfassende Analyse aller Metaboliten ab. Das ultimative Ziel ist die Identifizierung und Quantifizierung der strukturell verschiedenen kleinen Moleküle in einer Zelle oder einem Organismus, wie zum Beispiel einer Pflanze. Metabolomics ist ein zunehmend wichtiges Werkzeug in der Pflanzenwissenschaft; es gestattet, die Stressantworten von Pflanzen zu beurteilen und Phänotypen zu charakterisieren. Des weiteren unterstützt Metabolomics die Pflanzenzüchtung und erlaubt generell ein besseres Verständnis von zellulären Systemen. Die Zielsetzung dieser Arbeit war es, gezielte auf Massenspektrometrie basierte Metabolomics Methoden zu entwickeln. Ein erster, breiter Metabolomics Ansatz beabsichtigt den Nachweis von Unterschieden in metabolischen Fingerabdrücken und wurde angewandt, um fungale Pathogenresistenz in Getreide zu charakterisieren. Ein zweiter, gezielter Metabolomics Ansatz wurde konzipiert, um definierte Metaboliten eines bestimmten Stoffwechselweges zu beobachten. Dieses Vorgehen wurde für die Analyse einer spezifischen Klasse von Metaboliten, den Flavonolen, in der Modellpflanze *Arabidopsis thaliana* verwendet.

Wir haben erfolgreich Pflanzenmetabolomics Workflows entwickelt für die Untersuchung des Resistenzgens *Lr34*, geschaffen durch kombinierte Bemühungen von Seiten der Pflanzenbiologie und der analytischen Chemie. Die metabolische Charakterisierung des Resistenzgens *Lr34* hat ermöglicht, ausgeprägte Unterschiede zwischen Kontroll- und *Lr34*-Pflanzen zu erkennen, dafür verantwortliche Metaboliten zu identifizieren und diese Metaboliten zur biologischen Interpretation in Stoffwechselwege einzubetten.
Die Einführung von dauerhaften Resistenzgenen in Nutzpflanzen ist eine wichtige Strategie, um Ertragsverluste durch Pilzerreger zu verhindern und die Ernährungssicherheit aufrecht zu erhalten. Das Resistenzgen *Lr34* von Weizen (*Triticum aestivum*) verleiht dauerhafte Resisenz gegen vier wichtige biotische Pilzkrankheiten: Braunrost, Gelbrost, Stängelrost und Mehltau. Während *Lr34* in Weizen nur bei erwachsenen Pflanzen Resistenz hervorbringt, ist *Lr34* funktionell übertragbar auf Gerste (*Hordeum vulgare*) und Reis (*Oryza sativa*), wo es im Setzlingstadium exprimiert wird, was die Gewinnung von Pflanzenmaterial innerhalb einer kürzeren Zeitspanne erlaubt.
Der molekulare Mechanismus der *Lr34* Resistenz, dessen Gen für ein ABC-Transportprotein codiert, ist noch nicht bekannt. Das übergeordnete Ziel des multidisziplinären Projektes war es, das Verständnis der molekularen Funktion und der

Abwehrreaktionen im Zusammenhang mit der dauerhaften Krankheitsresistenz in Getreide zu verbessern. Die metabolische Charakterisierung des Krankheitsresistenzgens *Lr34* wurde in dieser Arbeit in zwei Schritten durchgeführt.

In einer ersten Reihe von Experimenten wurden die metabolischen Fingerabdrücke von Kontrollpflanzen und transgenen *Lr34* Gerstenpflanzen mittels Ultrahochleistungs-Flüssigkeitschromatographie (UHPLC)-Massenspektrometrie (MS) basierter Analyse verglichen. Dabei richtete sich der Fokus auf Sekundärmetaboliten. Die Pflanzen wurden auf hydroponischer Lösung gezüchtet, um die Wachstumsbedingungen zu kontrollieren, basierend auf der Hypothese, damit weniger biologische Variabilität zu induzieren. Signifikant veränderte metabolische Komponenten wurden für die Annotation ausgewählt, und zwei Gruppen von hochregulierten Metaboliten wurden identifiziert.

Basierend auf Daten aus Hochauflösender Massenspektrometrie (HR-MS) und Tandem-Massenspektrometrie (MS/MS) wurden Hordatine, gerste-spezifische Hydroxyzimtsäure-amine, und Cyanoglykosidderivate annotiert. Die biologische Funktion von Cyanoglykosiden in Gerste wird noch nicht vollständig verstanden, aber Hordatine sind bekanntlicherweise an Abwehrreaktionen beteiligt und sind nach Pilzpathogenbefall induziert. Wahrscheinlich stellen diese beiden Metabolitenklassen indirekten Antworten dar, welche durch das Substrat vom LR34-Transporter verursacht werden, weil nur eine partielle Heraufregulierung beobachtet wurde. Vermutlich werden sie durch den starken Blattspitze-Nekrose Phänotyp in Gerste beeinflusst.



Schema 1: Klassen von Cyanoglycoside Derivatives und Hordatine, rauf regulierte Metaboliten in *Lr34* Gerste, veranschaulicht mit Epidermin beziehungsweise Hordatine A.

Eine zweite Versuchsreihe konzentrierte sich auf die Untersuchung der Stoffwechselantworten von *Lr34*. Das Ziel war, die Robustheit des Pflanzenwachstums zu berücksichtigen, *Lr34* in verschiedenen Getreidearten zu vergleichen und die Auswirkungen des Blattspitze-Nekrose Phänotyps zu umgehen, indem Gerste- und

Reisgenotypen mit niedrigen und hohen Expressionsniveaus von *Lr34* untersucht wurden. Aus diesem Grund wurde das Design der Untersuchung auf verschiedene Getreidearten erweitert. Dabei wurden verschiedene Wachstumsbedingungen einschliesslich hydroponischer, mock- und Pilzpathogen infizierter Experimente untersucht und weitere Klassen von Metaboliten analysiert. Zusätzlich wurden metabolischen Profile von im Feld gewachsen *Lr34* Weizen analysiert, um das Resistenzgen auch unter natürlichen Bedingungen zu untersuchen. Eine breite Palette von Substanzen, einschließlich primären und sekundären Metaboliten, Lipiden und ausgewählten Pflanzenhormonen wurden durch eine Kombination von UHPLC- und Gaschromatographie-MS-basierten Methoden analysiert.

Die wichtigsten Unterschiede im Stoffwechsel wurden in hochexprimierenden *Lr34* Gersten gefunden. In Reis und Weizen wurde der Glyoxylatzyklus induziert und die entsprechenden Metaboliten Isocitrat herunter- und Succinat hochreguliert. Ein möglicher metabolischer Ausgang des Glyoxylat-Zyklus und der nachfolgenden Gluconeogenese ist Glucose. Sie ist möglicherweise als Signalmolekül oder als Nährstoff involviert in die Metabolitenantwort von *Lr34* auf die biotrophen Pilze, welche sich von Zucker aus der lebenden Pflanze ernähren. Weiter wurde eine Gruppe von herunterregulierten Sekundärmetaboliten basierend auf HR-MS und MS/MS-Daten putativ als *C*-glycosylierte Flavone in *Lr34* Reis und Gerste annotiert. Flavonoide sind bekannt als defensive Verbindungen. Ihre Biosynthese war in *Lr34* Gerste auf mRNA-Ebene induziert, was den Ergebnissen von Metabolomics entgegen steht.

Insgesamt ergab diese Arbeit Einblicke in Stoffwechselwege, welche an der dauerhaften Multipathogen-Resistenz *Lr34* beteiligt sind. Sie trug zudem bei zu einem besseren Verständnis der metabolischen Antwort von Pilzpathogen-Resistenz in Getreide Gerste, Reis und Weizen. Generell wurde die biologische Interpretation dadurch erschwert, dass die direkt vom LR34-Transporter verursachten und die indirekt durch den starke Blattspitze-Nekrose Phänotyp in Gerste verursachten Effekte nicht eindeutig unterschieden werden konnten.

Im zweiten Projekt entwickelten und verbesserten wir eine UHPLC-MS-Methode für die gezielte Analyse von Flavonolen, um ihre biologischen Funktionen in Bezug auf Pflanzenwachstum zu untersuchen. Das beinhaltete die Annotation von bereits gefundenen und neuen Flavonolen.

Flavonoide sind eine wichtige Gruppe von Sekundärmetaboliten mit einer Vielzahl von Funktionen (z.B. UV-Schutz, Hormontransport, Zellentwicklung, Genexpression). Die

Modellpflanze *Arabidopsis thaliana* beinhaltet eine Untergruppe der Flavonoide, die Flavonole, welche zusammengesetzt sind aus Kaempferol, Quercetin oder Isorhamnetin, modifiziert durch Glucose- und Rhamnoseeinheiten. LC-MS ist die übliche Methode der Wahl, um Flavonole in komplexen Pflanzenextrakten zu charakterisieren. Eine zuvor entwickelte Methode zur gezielten Analyse von Flavonolen in Extrakten von *A. thaliana* wurde verbessert hinsichtlich Selektivität, Empfindlichkeit, Massengenauigkeit und Geschwindigkeit. Die Flavonolannotationen wurden evaluiert auf Basis von biologischen Mutanten, auf der Berechnung chemischer Formeln aus HR-MS Daten sowie basierend auf MS/MS-Spektren.

Flavonolprofile, welche mit der neuen Methode analysiert wurden, wurden erfolgreich angewendet, um die Rolle der Flavonole im Zusammenhang mit der Modifikation des Auxin-Transports, dem Zellwachstum und der Pflanzenentwicklung zu untersuchen. Somit ermöglicht die verbesserte Methode eine genauere, gezieltere Analyse der Flavonole in Pflanzen und stellt ein vielseitiges Werkzeug dar, um ihre komplexen und vielfältigen Funktionen zu studieren.

— Chapter 1 —

General introduction

General introduction

1 Plant metabolomics

Metabolomics is the systematic study of the total content of low molecular weight components (< 1500 Da) of a biological system; a cell, tissue, biofluid, cell culture or organism.

Metabolites are the end products of cellular regulatory processes under given genetic, nutritional or environmental conditions, and thus metabolomics provides the most "functional" information of the 'omics' technologies [1] and captures a snapshot of the biological system under investigation [2]. The ultimate goal is the comprehensive identification and quantification of all metabolites in a plant, organ or tissue and the application of plant metabolomics often involves comparison of multiple conditions arising from experimental biotic or abiotic perturbations or from natural fluctuations [3]. In plants, alterations of metabolite levels are caused by the abiotic environment, by genetic modification or by the interactions of the plant with pathogens and pests [4]. A single plant species was estimated to have a comparable number of metabolites and genes [5], and around 1–200'000 are expected in the whole plant kingdom [6].

While genome, transcriptome and proteome are correlated with a linear relationship and share predictable structures, the metabolome cannot be predicted from proteome [7]. The especially high chemical diversity in the plant kingdom [8] and the associated contrasting physicochemical properties of the plant metabolome comprise a huge analytical challenge. Currently, no single analytical technique is capable of detecting all the structurally diverse metabolites within a given biological sample; multiple analytical platforms are required in order to cover a wide range of metabolites [3]. An additional challenge is the vast variation in relative abundances of the metabolites with a huge dynamic range (fmol – mmol) [9].

Essential metabolites sustaining normal growth, development and reproduction of a plant are termed "primary metabolites" [10], and include amino- and organic acids, lipids and carbohydrates [11]. Non-essential metabolites with additional functions in a given environment (e.g. defense, signaling or pollinator attraction [12]) are called "secondary metabolites" and consist of polyphenols, alkaloids, terpenes, polyketides and hormones [11]. While primary metabolites are highly conserved among the plant kingdom, secondary metabolites display a high diversity. The high secondary metabolite diversity was influenced by local or whole genome duplication in plants [13].

Metabolomics is a promising area in plant research, and a broad variety of applications have emerged in recent years. In combination with other systems biology approaches, plant metabolomics will be of extensive value for a better understanding of cellular systems and the regulatory mechanisms of metabolism [14]. For example, the discovery of novel pathways for the biosynthesis of natural products, the annotation of metabolism-related genes and the identification of new gene functions is enabled by combining metabolomics and genomics [7]. Other typical applications of plant metabolomics are the characterization of mutants and phenotypes, or the assessment of plant stress responses and interactions with the environment [15]. For example, metabolomics was utilized to investigate plant response to abiotic stress such as draught [16, 17] and heat [18] in wheat, nitrogen deficiency [19], salt stress [20] and boron toxicity [21] in barley, or exposure to Cr [22] and Cd/Cu [23] in rice.

Furthermore, metabolomics is emerging as a powerful tool for plant breeding. Metabolomic approaches offer opportunities for the identification of resistant biomarker metabolites that combine effect from genetic background and environmental factors [24], they enable the selection of elite germplasm for breeding [7], allow the amelioration of seed nutritional quality traits [14] or the development of cultivars with high levels of secondary metabolite neutraceuticals [25]. The technology also plays a key role in the public acceptance of genetically modified crops, as it provides necessary data for risk management [26]. Together with current developments in genomics and proteomics, metabolomics can contribute to quality and safety of future global food supply [4].

Two complementary strategies are widely used in current metabolomics: targeted and untargeted analyses. Targeted metabolomics reflects the traditional approach of monitoring a predefined, known set of metabolites usually related to a particular metabolic pathway and often includes quantification with authentic reference standards [11].

Contrasting, untargeted metabolomics potentially seeks the detection of the entire metabolic space and uses relative quantification [3]. Within untargeted metabolomics, literature differentiates between metabolic fingerprinting and profiling. Metabolic fingerprinting is a global and data-driven high-throughput approach with the primary aim of sample comparison and discrimination analysis [27]; the focus is on pattern comparison to highlight differences between treatment groups. Metabolic profiling is a qualitative and usually quantitative analysis of different classes of metabolites, resulting in comprehensive lists of metabolites in a sample [28]. Often, various analytical approaches are combined to achieve a comprehensive picture on primary and secondary metabolites.

The general concept of metabolomics matured in recent years [7] with an increasing number of publications on the topic (Figure 1). One of the remaining bottlenecks remains the annotation of metabolomic signals, especially in untargeted approaches [26].

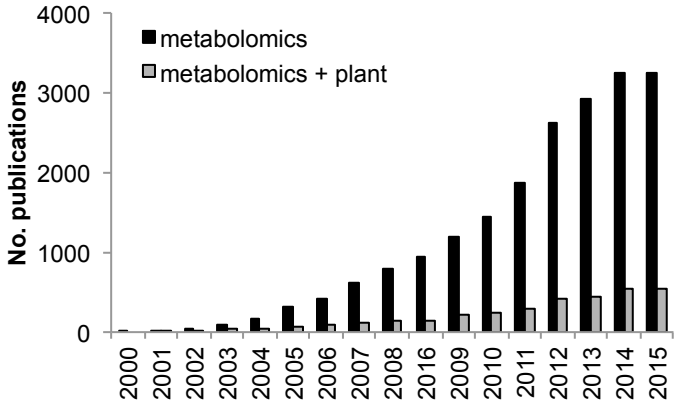


Figure 1: Bibliographic search in Chemical Abstract Service (CAS) *SciFinder* for keyword "metabolomics" (black) refined with keyword "plant" (grey).

Metabolomics informs about metabolites' amounts and identity, but only provides a static snapshot of a biological system. Sophisticated approaches additionally investigate the subcellular distribution [29, 30], the localization or the dynamic metabolic fluxes of metabolites. Together, these recent technical advances foster the functional and mechanistic understanding of cellular processes within the whole organism [29].

Mass spectral imaging is a technique that allows the spatially resolved observation of metabolites within a tissue or a cell. Therefore, a laser or ion beam is screened over a sample surface, desorbing and ionizing molecules that are transferred into a mass spectrometer. By selecting a specific mass-to-charge ratio (*m/z*), the signal of a metabolite of interest can be monitored across the image [31]. A number of ionization techniques are used for imaging, e.g. secondary ion mass spectrometry (SIMS) [32], desorption electrospray (DESI) [33] under atmospheric pressure or matrix-assisted laser desorption (MALDI) [34, 35], and the spatial resolution ranges from 0.2 to 100 µm [36]. Although sample preparation for plant tissue is not yet as far advanced as for animal tissue [31], mass spectral imaging is emerging in plant science. Examples are the observation of glucosinolates [37] and flavonoids [38] in *Arabidopsis* flowers, poly- and oligosaccharides in wheat seeds, or anthocyanins in rice pericarps [39]. Another development is the direct metabolomics analysis of single (plant) cells by MS [40, 41].

Metabolic flux analysis investigates the velocity of molecular reactions *in vivo* by observing metabolites labeled with stable isotopes such as ^{13}C or ^{2}H, but data analysis of non-

stationary labeling experiments is substantially complex [12]. Notwithstanding, flux analysis enables to unravel sites and mechanisms of metabolic regulation, therefore expanding our understanding of cellular metabolic interplay [12]. As a further technological extension, a combination of special resolution and dynamics of metabolites had recently been described as metabolic flux imaging [42].

2 Metabolomics workflow

To achieve the goals of the different targeted and untargeted metabolic approaches, it is key to have a functional combination of comprehensiveness, analytical precision and sample throughput [27]. Despite the broad applications of metabolomics, the workflow for metabolomic experiments remains generally very similar. This chapter aims to give a general introduction to the distinct steps of the metabolomics workflow and to discuss according procedures and challenges. An overview of the major processes for generating metabolomic data is given in Figure 2.

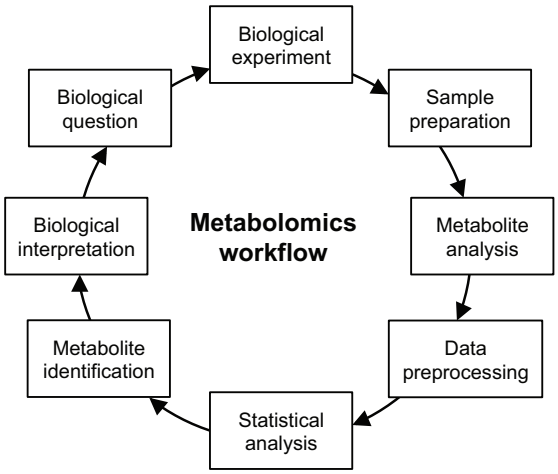


Figure 2: General workflow of a metabolomics experiment.

Starting point of the metabolomics workflow is a biological question, according to which a biological experiment is designed. Sample preparation and metabolite analysis are followed by data preprocessing, statistical analysis and metabolite identification to finally extract biological insights.

2.1 Biological question and biological experiment

As mentioned above, a vast range of biological questions can be addressed by metabolomic approaches to investigate changes — which may be caused by genetics, nutrition or environment — in a plant compared to a control.
The experimental design should comprise a sufficient number of replicates to account for biological and experimental variations. Usually, biological replicates are plants from independent sources of the same genotype, grown under identical conditions [43]. Three to six biological replicates were recommended for each experimental condition [43, 44]. Pooling of the individual plants was only recommended for insufficient sample amount [44].
In order to study the biological questions, considerations about possible undesired perturbations in the biological experiments need to be addressed. In plant experiments, this involves a rational selection of growth conditions, environmental influences, developmental stage and treatment. To address fluctuations in the environment, controlled growth conditions are preferable [45].

2.2 Sample preparation

Sample preparation is a fundamental and critical step with important consequences for the accuracy of later results. It should be unselective and reproducible and must prepare the sample appropriately for the subsequent data acquisition technique [46]. Depending on the plant / organ, different harvesting, homogenization and extraction techniques are applied. In contrast to data acquisition, which is usually automated, the sample preparation is mostly done manually [45]. It is important to keep the procedure fast and as simple as possible to achieve highest reproducibility, and the use of blank samples is indispensable to identify and eliminate possible laboratory contaminations [44].

2.2.1 Harvesting

Plant metabolites, both primary and secondary, are highly influenced by the environment of the plant, especially in their natural inhabitant [47]. Under controlled conditions such as provided in a greenhouse, the environmental influence is less pronounced. Important factors for the selection of harvesting conditions are the developmental stage of a plant and the time of harvest [45], as variations of metabolite levels such as flavonoids or alkaloids have been observed throughout the day [47].
Quick harvesting procedures are aspired to inactivate metabolic turnover as fast as possible after harvesting. Indeed, wounded plants react within seconds by producing

signal molecules such as jasmonic acid [48]. Immediate inhibition of enzymatic activity is usually achieved by flash-freezing the sample in liquid nitrogen or even by freeze-clamping tissues between precooled metal plates. Metabolite contents vary between plant organs and even differ within leaves [49], therefore separation of organs is recommended.

2.2.2 Homogenization

Homogenization of plant material into small particle aims at improving the extraction process. Homogenization can be done manually by mortar and pestle, or automatically by a ball mill or analytical mill. It is of high relevance to keep the plant material frozen during the homogenization procedure; thawed samples need to be discarded [50]. This is equally valid for plant powder weighing, which needs to be done as quickly as possible.

For some analyses, e.g. nuclear magnetic resonance (NMR) spectroscopy, the samples need to be dried. In most cases, freeze-drying is the gentlest drying method. However, even during lyophilization some metabolites may undergo degradation, e.g. succinate or choline [51].

2.2.3 Extraction

Extraction methods should be simple and fast. Due to the high diversity of plant metabolites comprising apolar (lipids, fatty acids, terpenoids), medium polar (secondary metabolites; alkaloids, flavonoids) and highly polar compounds (primary metabolites; sugars, amino acids), no solvent can concomitantly extract all metabolites [45]. Apart from solvent characteristics, other factors such as the ratio of solvent to sample, the duration or temperature of the extraction influence the extraction efficiency. The choice of the solvent encompasses a very strong bias to the view on the metabolome [45], but it needs to be compatible with the data acquisition technique. To broaden the coverage, mixtures of solvents [52, 53] or repetitive extraction with different solvents are applied [45, 54]. Methods to accelerate metabolite extraction involve shaking, ultrasonic extraction, microwave-assisted or pressurized solvent extraction [47].

2.3 Metabolite analysis

Various analytical methods have found applications in plant metabolomics, including mass spectrometry (MS) – either coupled to separation techniques or as direct infusion mass spectrometry (DIMS) – or NMR spectroscopy [1]. The most prominent separation methods are liquid and gas chromatography (LC and GC, respectively), but also capillary electrophoresis has been applied. Recent technical advances in metabolomics involve

multidimensional chromatographic separations (e.g. LCxLC, GCxGC), mass spectral imaging or coupling of LC-MS with solid-phase extraction and NMR [11].

2.3.1 Mass spectrometry

Mass spectrometry-based methods are highly sensitive with detection limits at femto to pico moles [47]. They allow relative quantitation and provide information about the masses of compounds.

A mass spectrometer is typically composed of three parts: an ion source, a mass analyzer and a detector. The ion source converts sample molecules to ions and the mass analyzer resolves these ions based on their *m/z* before a detector records the signals [55]. Most commonly for liquid chromatography, atmospheric pressure ionization methods with a strong prevalence for electrospray ionization (ESI) and less frequent atmospheric pressure chemical ionization (APCI) are used [47]. ESI results mainly in intact ionized molecules and is therefore considered a “soft ionization” method. Atmospheric pressure ionization methods allow the detection of ions in negative and positive ionization mode, which are usually combined in metabolomics. As some metabolites are observed exclusively either as protonated $[M+H]^+$ or as deprotonated $[M–H]^–$ molecules [56], the combination of both ionization modes consequently broadens the range of detectable compounds [57].

The major ionization technique used in combination of GC and MS is the rather harsh electron impact (EI) ionization, which provides highly reproducible fingerprint spectra of metabolites [57]. While ESI primarily produces single intact ionized molecules and a few adduct ions, EI generally results in a variety of fragments.

Several types of mass analyzers are available that differ in their basic characteristics regarding mass accuracy, resolving power and acquisition speed. Mass accuracy is defined as the relative difference between the experimental *m/z* value and the effective *m/z* value of an analyte. Another basic characterization for mass analyzers is the ability to resolve peaks in mass spectra. The resolving power of a mass analyzer is defined as the *m/z* value of a particular mass peak divided by its full width at half maximum and is always determined for a particular *m/z* value [58]. Examples for mass analyzers with low resolution are quadrupole and ion trap analyzers. High-resolution analyzers with resolving power 10’000 – 100’000 are time-of-flight (TOF), Orbitrap, sector field instruments or Fourier-transform ion cyclotron resonance analyzers, and they are also capable of high mass accuracy measurement (< 5 ppm). Because ESI is a soft ionization technique, it provides limited structural information about the analyte. To gain more structural information or to quantify metabolites, tandem mass spectra (MS/MS) are recorded. In the

process of collision-induced dissociation (CID), precursor ions are collided with inert gas (e.g. N_2, He or Ar) to produce characteristic fragments. Tandem-in-space instruments combine two mass analyzers with a collision cell in between, allowing the spatial separation of precursor and fragment ions (e.g. triple quadrupole or Q-TOF instruments). Tandem-in-time instruments such as ion traps are capable to perform multiple stages of MS (MS^n). High-resolution instruments can also be used without chromatography by flow- or direct infusion [59] to achieve high sample throughput. Accurate mass determination with precision in the ppm range allows to massively decrease the ambiguity of ion identification [60]. Direct infusion mass spectrometry suffers from ion suppression, arising from competitive ionization of analytes co-eluting with matrix components, and isomers cannot be distinguished [11].

2.3.2 Nuclear magnetic resonance

Nuclear magnetic resonance is often applied for metabolic fingerprinting of all hydrogen-bearing metabolites. Particularly, ^{1}H-NMR analysis of the non-separated samples are performed and analyzed by multivariate or pattern recognition techniques such as principal component analysis [61]. A major advantage of NMR analysis is the simple quantitation, whereas the major disadvantage is the very low sensitivity compared to mass spectrometry. The sensitivity of MS is several orders of magnitude lower compared to NMR [11]. Due to recent technical advances (microprobe, cryoprobe, improvement in pulse sequence), no more than 10 – 50 mg of biomass is required for NMR profiles. Signal overlap in ^{1}H spectra can be resolved by two-dimensional techniques [62] revealing direct structural information of the individual metabolites (typically up to 150), but at the cost of longer acquisition times. NMR-based metabolomics has been applied for example to the analysis of the equivalence of genetically modified plants [63]. Combined with prior targeted MS micro-isolation [64] or with solid-phase extraction trapping of metabolites [65], NMR has proven especially helpful in structural identification of novel metabolites.

2.3.3 Separation

Hyphenation of mass spectrometry with chromatography leads to higher analytical sensitivities but enhances the analysis time.

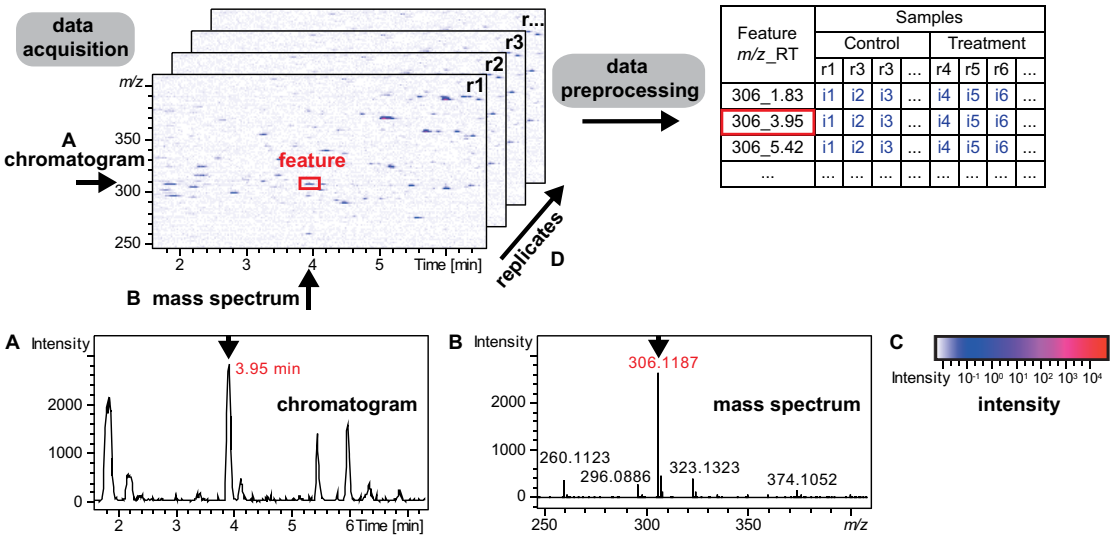


Figure 3: Three-dimensional data from LC-MS data acquisition: chromatogram at *m/z* 306.1 ± 2.0 (A), mass spectrum at 3.95 min (B), intensity color code (C). After data preprocessing, data are summarized in a table comprising features characterized by a specific *m/z*-RT pair in each row and the respective intensities of the different replicates in each column.

For each elution scan point, a mass spectrum is recorded, generating three-dimensional data (Figure 3). A certain signal defined by a given *m/z* ratio and retention time (RT) is called feature, e.g. *m/z* 306.1187 at 3.94 min.

In metabolomics, mainly liquid chromatography (LC) and gas chromatography (GC) are used, besides niche applications of capillary electrophoresis, especially for charged metabolites [11]. LC is suitable for relatively polar metabolites with a broad range of masses, and reversed phase (RP) chromatography had widely been applied for (semi)polar secondary metabolites and lipids [66]. Advances in technology, such as ultra-high performance liquid chromatography (UHPLC), allow the detection of up to several hundred metabolites in a complex plant extract in one chromatographic run [50]. UHPLC columns are filled with < 2 µm particles and have a high chromatographic resolution and separation efficiency resulting in narrow peaks (peak width 3 – 10 s), which increase the peak capacity and allow short chromatographic runs. To obtain sufficient data points, mass analyzers with high acquisition speed (e.g. TOF) are required. The better chromatographic resolution reduces ion suppression during electrospray ionization and in improves signal-to-noise ratio, hence leads to increased sensitivity [67]. The main disadvantage of the small particles is their high backpressure rising up to 1300 bar [58].

GC-MS is applied to hydrophobic, heat stable and volatile compounds with low molecular weight (< 350 Da), such as hydrocarbons, esters, and metabolites with reduced polarity

[47]. With an additional derivatization to increase the volatility and thermal stability of the compounds, primary metabolites such as amino acids, organic acids, sugars and sugar alcohols can be analyzed [45]. Modern capillary GC columns provide extreme peak capacity for high-resolution separation of metabolic extracts. GC-MS provides reproducible and accurate measurements of fragmentation patterns and many libraries have been built to facilitate the identification of compounds [56]. Metabolomics data generated by GC-MS are exceedingly complicated because of the complex signal patterns that result from fragmentation during EI ionization. Metabolites may result in different chemical derivatives, e.g. incompletely derivatized analytes or analytes containing multiple trimethylsilyl groups. Taken together, GC-MS based metabolomics data are intrinsically more complex than LC-MS data, which contain a multitude of features – caused by derivatization and fragmentation – corresponding to a single metabolite [68].

2.3.4 Methodological considerations

Metabolomes are complex and include a wide range of metabolites of different polarity, acidity or basicity and reactivity [4]. To cover the metabolome as comprehensively as possible, multi-parallel approaches need to be combined applying complementary extraction, separation and detection technologies [9]. Hot polar solvents were found as most efficient in extracting both hydrophilic and hydrophobic metabolites simultaneously [53]; on the other hand, elevated temperatures may well destroy reactive metabolites.

The combination of orthogonal GC- and LC- based analyses offers a complementary approach to detect metabolites with different physico-chemical properties, thus broadening the metabolic coverage [11].

Most of the steps in the metabolomics workflow introduce a certain bias towards the finally detected metabolites, namely time and procedure of sample collection, extraction conditions but also dilution and storage effects of the sample must be considered. Further limitations arise from the types and numbers of analytical methods used, ion suppression (LC-MS), derivatization (GC-MS), sensitivity of the assay, range of reliable response and the ability to allow at least for semi-quantitative comparison [69].

However, a pragmatic approach must be taken to facilitate not only comprehensive metabolome coverage, but also to allow a sufficiently high sample throughput [67]. To minimize the influence of uncontrolled variables, spatiotemporal randomization of biological replicates is aspired optimally not only for instrumental analysis but also throughout the biological experiments and the sample preparation steps [43].

2.4 Data preprocessing

After metabolite analysis, the resultant raw data need to be converted into an appropriate form for the subsequent statistical analysis. The aim is to collapse the three-dimensional analytical data (retention time / *m/z* / intensity) into a two-dimensional peak table for further statistical analysis (Figure 3) [67]. Within the table, each row contains a specific feature and each column the intensities of the specific feature in all replicate samples. A metabolite causes at least one feature, but might also give raise to additional adduct / fragment signals. Preprocessing is performed for a whole experiment involving all samples, to ensure that a single feature correlates between different replicates and different treatment groups (e.g. control and resistant plants).

Data preprocessing for LC-based raw data usually includes detection of signal peaks (peak picking), alignment, normalization, and generation of a data matrix including all peaks of a given sample set [66].

GC-EI-MS data are intrinsically more complex than LC-ESI-MS data: a single metabolite can yields several analytes due to incomplete derivatization, and each analyte produces numerous fragments. As a result, metabolite identification is usually performed in advance to statistical analysis to reduce the amount of data. Accordingly, preprocessing of GC-MS raw data involves the following steps: retention time alignment, deconvolution, identification and quantification based on a single fragment. Data processing e.g. by *TagFinder* software [70] allows the standardized conversion of GC-MS information to comprehensive numerical data matrices and was developed for the high throughput quantitative data analysis of GC-MS based metabolite profiling experiments.

Peak picking involves the detection of peaks across the spectrum and the integration of the according peak area to provide a list of two-dimensional signals, so-called features – defined by a unique *m/z* – retention time combination [71]. Feature detection is a crucial step, which needs to be reliable and sensitive; the algorithm must be able to detect low-intensity features but it should at the same time be able to ignore feature-like signals such as chemical noise [72].

The simplest strategy for peak picking is to apply an intensity threshold; constraints on the peak shapes in the chromatographic direction can refine the peak detection. Different possibilities are based on the examination of extracted ion chromatograms (EIC) or on the application of a model-fitting procedure involving isotope detection [56].

Various software for feature detection is available commercially (e.g. *MarkerLynx*, Waters), but also freely (e.g. *MetAlign* [73], open-source *XCMS* [74], *MZmine* [75]). An algorithm suited for LC-MS metabolomics data is *centWave*, which is included in the *XCMS* data processing tool. The *centWave* algorithm sensitively detects potentially interesting mass traces, termed regions of interest, followed by an extensive chromatographic analysis able to detect chromatographic peaks with varying width by continuous wavelet transform. It has been shown to be highly sensitive, as baseline and noise are estimated locally for each peak [72].

Importantly, a specific feature must always correspond to the same signal within different replicates. Therefore, a mass calibration is done in *m/z* dimension and a retention time alignment in the RT dimension.

Retention time variability in LC-MS is affected by physical uncertainties (e.g. changes in temperature, pH, pump fluctuations) and drifts in the retention time dimension are often non-uniform along the chromatographic run. In order to combine chromatographic runs of different samples and to ensure a one-to-one correspondence between features, a proper alignment of retention times is necessary [56]. In *XCMS*, RT distortions are compensated by a de-warping algorithm, which is based on a regression between landmark features used to improve grouping of peaks across samples. Usually a pooled quality control sample is applied as reference for alignment with the ordered bijective interpolated warping (*OBI-Warp* [76]) algorithm that corrects the retention time in all samples simultaneously in a single step [74].

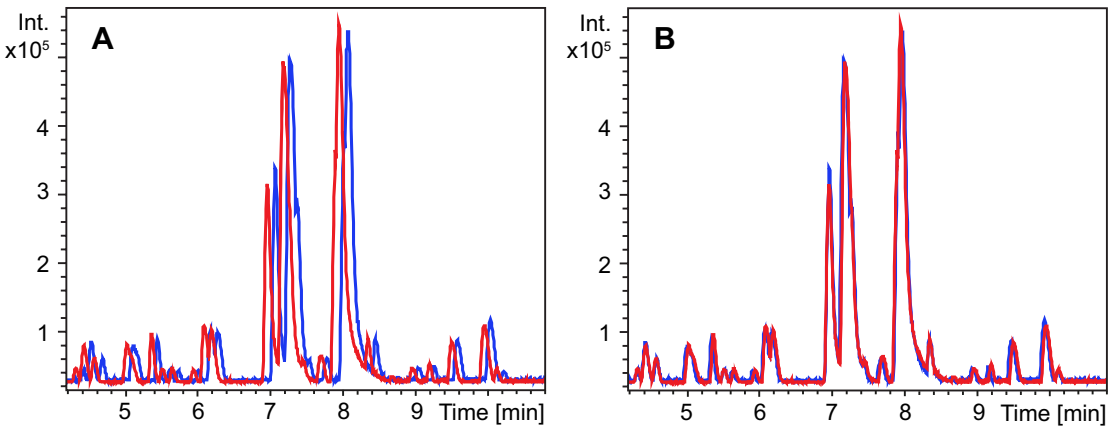


Figure 4: Two chromatograms before (A) and after retention time alignment (B).

In order to extract the most relevant features from the raw data, normalization of peak intensities is performed to remove undesired systemic bias and deviation in signal intensities due to analytical variation based on sample preparation or instrumental drifts [77, 78], Normalization can be performed sample-wise or feature-wise. Sample-wise normalization corrects for variation in sample preparation and analysis and involves correction with regard to internal standard compounds, dry weight, a constant sum or a reference sample. Feature-wise normalizations adjust the magnitude of features, e.g. to correct for concentrations varying by orders of magnitude in typical metabolomics samples. Frequently applied feature-wise normalizations are transformations, centering and scaling. Data transformations correct for non-symmetric variation by nonlinear conversions, e.g. by log or power transformation [79]. In centering, signals are converted to fluctuate around zero instead of around the mean. Scaling methods divide each variable by an individual scaling factor, either based on data dispersion (e.g. standard deviation) or on size measure (e.g. mean).

Following normalization, features are filtered to select the robust ones. Common filtering strategies comprise criteria for background noise level, intensity, variance, average signal or consistency within a treatment group.

2.5 Statistical analysis

In essence, statistical analysis aims at determining features that are relatively different between metabolic profiles of two or more distinct treatment groups to infer a biological relationship [80]. For example, a group of resistant plants is overall compared to a group of control plants, and differences are evaluated for variations within individual replicates.

A typical characteristic of metabolomics datasets is the large number of simultaneously observed variables (features) compared to the number of observations (biological replicates) [80]. Consequently, metabolomics data sets are multivariate, and multivariate analysis methods are applied to identify biological relevant spectral feature by consideration of the combinatorial effect of multiple variables. The probably most widely used statistical tool in metabolomic studies is principal component analysis (PCA), an unsupervised method that does not require knowledge about the treatment group associations of the samples [55]. PCA comprises an orthogonal transformation of the two-dimensional matrix, containing all features of a metabolomics approach, into a new coordinate system that uncovers hidden internal structures by building principal components that describe the maximum variance of the data [56]. Visualization is achieved by plotting the samples in the principal component space, in so called score plots

(Figure 5). The two-dimensional matrix is collapsed into a new coordinate system, and replicates of different treatment groups are visualized as single points. The separation of the different treatment groups in a PCA score plot indicates the potential discriminating power of the metabolic features in the two dimensional matrix [81]. A corresponding PCA loadings plot can reveal the contribution of individual metabolite features to each principal component. Alternatively, supervised methods require prior knowledge about the treatment group associations of the replicates, and a popular method is *partial least squares projection to latent structures* [80].

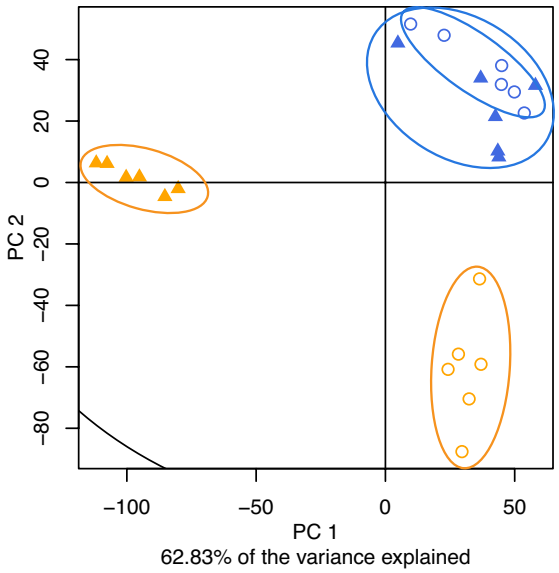


Figure 5: Example of a PCA score plot based on a two-dimensional matrix of LC-(–)-ESI-MS secondary metabolite features of pathogen-infected barley. Each individual circle or triangle corresponds to a replicate sample. The circles and triangles of the orange treatment groups were clearly separated, whereas no group separation was obtained for the blue circles and triangles. The first two principal components PC1 and PC2 together explained 62.83% of the total variance present in the matrix.

Univariate methods assess the statistical significance of each feature independently. Parametric tests are used to compare different treatment groups, containing a number of replicates, and include for example the Student's *t*-test or the analysis of variance (ANOVA) [82]. The tests observe the probability of the null-hypothesis proposing that the difference is only due to change and that there is no relationship between the treatment groups. A low probability value describes that the feature difference is unlikely due to chance. As a common significance threshold, features with probability values (*p*-values) < 0.05 are considered as significantly different.

Based on the high number of simultaneously analyzed features in metabolomics data sets, the probability of finding statistical differences by chance is high and even a *p*-value of 0.05 lead to many false discovery results [83]. This problem is called *multiple hypothesis testing problem* and can be addressed by multiple testing methods that adjust the *p*-values. An example of a multiple hypothesis testing method is the stringent Bonferroni correction, involving the multiplication of the number of comparisons with the *p*-values [84]. A less conservative approach, named after Benjamini and Hochberg, is based on minimizing the false discovery rate [85].

2.5.1 Data processing tools

Metabolite profiling by GC-MS depends on chemical derivatization, giving rise to one or more chemical derivatives per metabolite. Further complication arises from EI-induced fragmentation. Consequently, a high complexity of features is observed. In order to reduce the complexity of GC-MS data, metabolite identification was done prior to two-dimensional matrix generation. Basically, features were clustered and the resulting fragmentation pattern matched with reference spectra form a database in order to identify the metabolites. Out of the observed cluster, a single specific fragment mass was selected for relative quantification. In a next step, statistical analysis was performed to highlight significantly different metabolites.

In contrast, LC-MS data were preprocessed to result in a two-dimensional matrix containing features of yet unknown metabolite origin, by either *Genedata* or *XCMS* software. Different statistical analysis tools were used to find significantly different features, namely *XCMS*, *MetaboLyzer*, *Feature filtering* and ANOVA (Table 1).

Table 1: Overview of data processing tools used in this thesis, described in the following section. Software names are marked with an asterisk.

Tool	GC-MS data	LC-MS data	Preprocessing	Statistical analysis
*TagFinder**	X		X	
*Genedata**		X	X	
*XCMS**		X	X	X
*MetaboLyzer**		X		X
Feature filtering		X		X
ANOVA				X

2.5.2 Genedata

Genedata Expressionist software is a commercial workflow-based solution for data processing of MS-based raw data, with the *Refiner MS* tool for preprocessing including noise reduction, *m/z* calibration, RT alignment and peak detection. Retention time alignment is built on a pairwise alignment based tree algorithm, computing the similarity of a pair of chromatograms by correlating all individual spectra in a maximum retention time window.

2.5.3 XCMS

XCMS [72] is a package part of *Bioconductor* [86], a large open-source software project for bioinformatics, written in the platform independent programming language *R* [87]. The *XCMS* package is designed for metabolomics data processing (both LC- and GC-MS), specifically for preprocessing including feature detection, non-linear retention time alignment, statistical methods and also visualization tools. Its cloud based online version *XCMS Online* [88] offers a web-based intuitive graphical interface to process untargeted metabolomic data. It offers a simple, automated process for peak detection and retention time alignment between samples, determines differentially abundant metabolites, performs multivariate statistical analysis, and assigns possible metabolite identity based on the *Metlin* [89] database.

2.5.4 MetaboLyzer

The automated metabolite analysis tool *MetaboLyzer* [81] that runs on *Linux* focuses on statistical analysis. Metabolomics matrices often include many missing values, because features with certain *m/z* and retention time are only present in part of the samples. To overcome the issue of sparse matrices, *MetaboLyzer* highlights the ions that are present in a certain proportion of the samples (e.g. over 90%, termed complete presence data). Then, applying simpler statistical technique for the complete presence data, it extracts the relevant features with different abundances between control and treatment group. Next, it assigns metabolites to features by scrutinizing four small-molecule databases (e.g. *HMDB* [90], *KEGG* [91], *Lipidmaps* [92], *BioCyc* [93]).

2.5.5 Feature Filtering

The *Feature filtering* workflow [78] involves univariate statistical data analysis based on scripts programmed in *R* [87, 94].

Typical discovery profiling experiments based on LC-MS detect routinely thousands of features in biological samples [12] with redundancy in form of isotopes, adducts, dimers or fragments. For each instrument, there is a technical minimum threshold for isolating parent ions for subsequent reliable MS/MS fragmentation. Pooled quality control samples are ideally prepared by combining aliquots of each individual sample entering the dataset and analyzed throughout the whole LC-MS injection list. To select features worth the effort of metabolite identification including MS/MS fragmentation, the *Feature filtering* procedure calculates the analytical variance of features, and removes features that have a larger variance in repetitively injected quality controls than in biological replicates. Further criteria of *Feature filtering* involve intensity, analytical variation, statistical interference and fold change [95].

2.5.6 TagFinder

The *TagFinder* software toolbox assists automated and manually supervised metabolite identification with final data conversion into numerical data matrices for subsequent statistical analysis. Retention time alignment is facilitated by internal reference compounds, which are added during sample preparation (e.g. *n*-alkanes or fatty acid methyl ester) and allow the conversion of retention times into Kovàts' retention time indices (RI) [96]. Besides fragmentation patterns, also these retention time indices are used for metabolite identification. For peak picking, each nominal mass trace is screened and mass spectra are deconvoluted into fragments from coeluting analytes. Closely coeluting fragments are grouped and binned across overlapping RI windows, and mass spectra are reconstructed for manual visual inspection. Matching with MS/RI reference libraries like the *Golm Metabolome Database* [97] suggests putative metabolite identifications that allow manually supervised peak annotation [44, 68]. Within clusters, single quantification masses are selected based on criteria of specificity and intensity for subsequent statistical analysis.

2.6 Metabolite identification

After statistical analysis, the relevant features need to be identified as the responsible metabolites, in order to interpret biological meaning. This process of identification of discriminant biomarkers in complex metabolic mixtures remains the biggest challenge in metabolomics. Metabolite identification is still very time consuming (weeks to months), and often only small numbers of metabolites (<20) can be identified [12]. Consequently, the

vast majority of detected features in LC-MS based metabolic profiling experiments remain largely unknown [98].

The *Metabolomics Standards Initiative* [99, 100] defined a hierarchical annotation system with four levels of metabolite identification: 1) identification using authentic reference compounds, 2) annotation by verification of MS and MS/MS spectra with analytical data of authentic reference compounds, 3) characterization by interpretation of MS and MS/MS spectra and 4) unknown. Secondary metabolites are hard to purify and expensive to synthesize, thus authentic standards of potentially present metabolites in plants are often not available [43], and identification level 1 remains exceptional. Alternatively, a combination of physico-chemical properties – including the feature's *m/z* ratio, relative isotopic abundance, fragmentation and chromatographic retention time – are used to identify a metabolite [12]. The general strategy for metabolite identification involves the search in databases for molecular properties. To assign the correct molecular formula of a metabolite, adducts and dimers generated by ESI need to be recognized. Furthermore, comparison of signals in both ionization modes may assist assignment [11]. Mass spectrometers with high mass accuracy and high resolution facilitate metabolite identification by separating overlapping mass signals and generating precise isotopic fine structures [101]. Within a mass range limited by the mass accuracy of the instrument, possible elemental compositions are calculated and refined by heuristic filters. Heuristic filters include e.g. the isotopic pattern fit, limitations on the number of elements, restrictions of ratios of H, N, O, P or S in relation to C, Lewis rules, or element ratio probabilities [102]. Optionally, tandem mass spectra (MS/MS) fragmentation can be used to refine molecular compositions of the parent ion (e.g. in *SmartFormula3D*, Bruker Daltonics). The list of putative molecular formulas can be searched in various compound databases (e.g. *SciFinder* [103], *ChEBI* [104, 105], *PubChem* [106], *Dictionary of Natural Compounds* [107] or the plant-specific *KNApSAcK* [108]). An alternative, promising technique involves labeling of plant tissue with stable isotopes like ^{13}C, ^{15}N or ^{34}S, resulting in labeled metabolites with nearly identical physico-chemical properties to their native analogues. Spectra of the non-labeled aliquots are compared with the labeled isotopologs, and elemental-specific mass shifts significantly improve molecular formula annotation [52, 109]. Stable isotope labeling also allows the separation of truly biological derived signals from contaminants [110].

To further improve the annotation, fragment spectra based on MS/MS or MS^n, where *n* is the number of product ion stages, are matched with mass spectral databases (e.g.

MassBank [111], *ReSpect* [112], *METLIN* [113], *HMDB* [90], *LIPID MAPS* [114, 115], *LipidBlast* [116]). In comparison to highly reproducible EI spectra, LC-ESI-MS/MS spectra are instrument specific and different analytical conditions or collision energies cause different fragmentation patterns [98]. To overcome this effect, collision energy ramps are applied to generate an average spectrum showing the precursor ion and its fragments. For example, *METLIN* stores high resolution MS/MS at four different collision energies [89], *MassBank* includes spectra obtained under five collision energies [98, 111]. Although fragmentation patterns from MS/MS spectra are often difficult to interpret [109], already partial agreements with database spectra can suggest substructure information. A tool that combines spectral database search (*MassBank*) with *in silico* fragmentation (*MetFrag*) is *MetFusion* [117], using pairwise chemical similarities to calculate an integrated score, thereby improving the identification power and accuracy compared to simple *in silico* fragmentation. Other bioinformatics-assisted methods apply e.g. pairing of LC-MS features with known substrate / product pairs of enzymatic reactions [118], ranking of metabolite identifications based on network structure or probabilistic annotation [119], or validation of candidates based on prediction of retention times [101, 120].

De novo structure determination by NMR even provides stereochemical information, but extensive purification of the metabolites from complex mixtures is required. Automated approaches such as indirect coupling through LC-solid phase extraction-NMR are efficient for identification of metabolites in complex extracts; LC peaks triggered by the detector (MS or UV) are automatically trapped on solid phase extraction cartridges, eluted with deuterated solvents and directly transferred into a NMR flow cell [11].

After metabolites of interest have been identified, the findings require validation in a more comprehensive context. Follow-up experiments often involve targeted metabolomics to quantify the identified metabolites. In addition, biochemical verification experiments should be performed to assess the biological validity in more detail.

2.7 Biological interpretation

The biological interpretation of metabolomic data is a very challenging task. After metabolites of interest have been identified, information about the potential metabolic reactions connecting them is required. Metabolic networks stored in databases (e.g. *KEGG*, *BioCyc*) often contain hundreds of reactions and hundreds of metabolites. The graph structure of the network can be used for mining metabolomics data in the context of metabolic networks and to provide clues on reactions. Several software tools allow mapping and visualization of a set of metabolites in graphical representations like

metabolic pathways, e.g. *KEGG*, *AraCyC*, *Plant Metabolic Network* [121], or *MetaCrop* [98]. Other software tools are known for enrichment analysis [122] and for metabolic modeling [123]. The web server *MetExplore* is dedicated to the analysis of genome scale metabolic networks that allows analyzing metabolomics data in context of networks [124], and provides access to *BioCyc*, a collection of pathway and genome databases [93] including further information about cell compartments, pathways, reactions, metabolites, enzymes, proteins and genes.

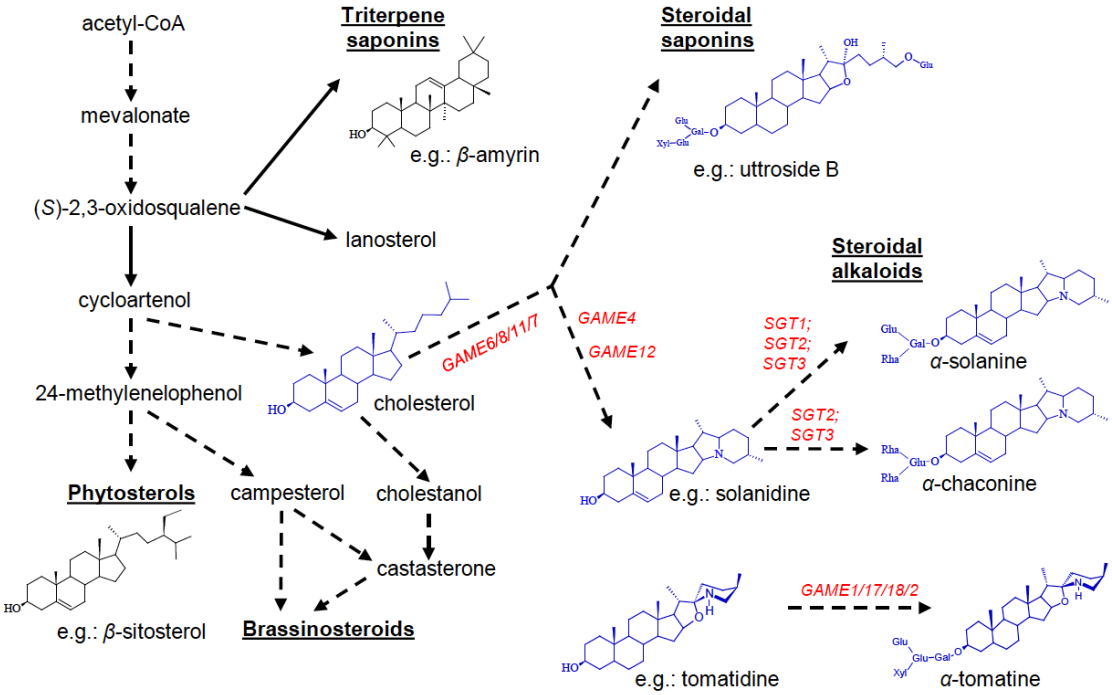


Figure 6: Biosynthesis of steroidal alkaloids and saponins in the tri-terpenoid biosynthetic pathway in Solanaceae plants. Dashed and solid arrows represent multiple or single enzymatic reactions in the pathway, respectively. Genes implicated in the steroidal glycoalkaloids pathway are marked in red [125].

Incorporation of the distinct metabolites in metabolic networks, together with transcriptomics and genomics data, represents a key issue to gain functional interpretation of the results. An example of a biosynthetic map of steroidal alkaloids, toxic compounds found in tomato and potato, which was deciphered by genomics, transcriptomics and network tools combined with metabolomics data, is shown Figure 6 [125].

3 References

[1] Wolfender, J. L., Rudaz, S., Choi, Y. H., Kim, H. K., *Curr. Med. Chem.,* **2013**, *20*, 8, 1056-1090

[2] Haug, K., Salek, R. M., Conesa, P., Hastings, J., de Matos, P., Rijnbeek, M., Mahendraker, T., Williams, M., Neumann, S., Rocca-Serra, P., Maguire, E., Gonzalez-Beltran, A., Sansone, S. A., Griffin, J. L., Steinbeck, C., *Nucleic Acids Res.,* **2013**, *41*, Database issue, D781-786

[3] Schuhmacher, R., Krska, R., Weckwerth, W., Goodacre, R., *Anal. Bioanal. Chem.,* **2013**, *405*, 15, 5003-5004

[4] Boggess, M. V., Lippolis, J. D., Hurkman, W. J., Fagerquist, C. K., Briggs, S. P., Gomes, A. V., Righetti, P. G., Bala, K., *J. Proteomics,* **2013**

[5] Verpoorte, R., Choi, Y. H., Mustafa, N. R., Kim, H. K., *Phytochem. Rev.,* **2008**, *7*, 525–537

[6] Allwood, J. W., Goodacre, R., *Phytochem. Anal,* **2010**, *21*, 1, 33-47

[7] Sumner, L. W., Lei, Z., Nikolau, B. J., Saito, K., *Nat. Prod. Rep.,* **2015**, *32*, 2, 212-229

[8] Tohge, T., Mettler, T., Arrivault, S., Carroll, A. J., Stitt, M., Fernie, A. R., *Front. Plant Sci.,* **2011**, *2*, 61

[9] Allwood, J. W., De Vos, R. C., Moing, A., Deborde, C., Erban, A., Kopka, J., Goodacre, R., Hall, R. D., *Methods Enzymol.,* **2011**, *500*, 299-336

[10] Tohge, T., de Souza, L. P., Fernie, A. R., *J. Chromatogr. B,* **2014**, *966*, 7-20

[11] Wolfender, J. L., Marti, G., Thomas, A., Bertrand, S., *J. Chromatogr. A,* **2015**, *1382*, 136-164

[12] Zamboni, N., Saghatelian, A., Patti, G. J., *Mol. Cell,* **2015**, *58*, 4, 699-706

[13] Scossa, F., Brotman, Y., de Abreu, E. L. F., Willmitzer, L., Nikoloski, Z., Tohge, T., Fernie, A. R., *Plant Sci.,* **2016**, *242*, 47-64

[14] Toubiana, D., Fait, A., in *Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield* (Eds.: Agrawal, G. K., Rakwal, R.), Springer Science+Business Media, Dordrecht, **2012**, pp. 453-475.

[15] Weckwerth, W., Kahl, G., *The Handbook of Plant Metabolomics*, Wiley-Blackwell, Weinheim, DE, **2013**

[16] Hill, C. B., Taylor, J. D., Edwards, J., Mather, D., Bacic, A., Langridge, P., Roessner, U., *Plant Physiol.,* **2013**, *162*, 3, 1266-1281

[17] Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., Roessner, U., *Molecular plant,* **2012**, *5*, 2, 418-429

[18] de Leonardis, A. M., Fragasso, M., Beleggia, R., Ficco, D. B., de Vita, P., Mastrangelo, A. M., *Int. J. Mol. Sci.,* **2015**, *16*, 12, 30382-30404

[19] Kovacik, J., Klejdus, B., Babula, P., Jarosova, M., *J. Plant Physiol.,* **2013**, *171*, 3-4, 260-268

[20] Widodo, H., Patterson, J. H., Newbigin, E., Tester, M., Bacic, A., Roessner, U., *J. Exp. Bot.,* **2009**, *60*, 14, 4089-4103

[21] Roessner, U., Patterson, J. H., Forbes, M. G., Fincher, G. B., Langridge, P., Bacic, A., *Plant Physiol.,* **2006**, *142*, 3, 1087-1101

[22] Dubey, S., Misra, P., Dwivedi, S., Chatterjee, S., Bag, S. K., Mantri, S., Asif, M. H., Rai, A., Kumar, S., Shri, M., Tripathi, P., Tripathi, R. D., Trivedi, P. K., Chakrabarty, D., Tuli, R., *BMC Genomics,* **2010**, *11*, 648

[23] Navarro-Reig, M., Jaumot, J., Garcia-Reiriz, A., Tauler, R., *Anal. Bioanal. Chem.,* **2015**, *407*, 29, 8835-8847

[24] Gauthier, L., Atanasova-Penichon, V., Chereau, S., Richard-Forget, F., *Int. J. Mol. Sci.,* **2015**, *16*, 10, 24839-24872

[25] Saxena, A., Cramer, C. S., *Adv. Crop Sci. Tech.,* **2013**, *1*, 2, 1-4

[26] Putri, S. P., Nakayama, Y., Matsuda, F., Uchikata, T., Kobayashi, S., Matsubara, A., Fukusaki, E., *J. Biosci. Bioeng.,* **2013**, *115*, 6, 579-589

[27] Hall, R. D., Hardy, N. W., in *Plant Metabolomics: Methods and Protocols, Vol. 860* (Eds.: Hall, R. D., Hardy, N. W.), **2012**, pp. 1-10.

[28] Tenenboim, H., Brotman, Y., *Trends Plant Sci.,* **2016**

[29] Krueger, S., Steinhauser, D., Lisec, J., Giavalisco, P., *Methods Mol. Biol.,* **2014**, *1062*, 575-596

[30] Klie, S., Krueger, S., Krall, L., Giavalisco, P., Flugge, U. I., Willmitzer, L., Steinhauser, D., *Front. Plant Sci.,* **2011**, *2*, 55

[31] Spengler, B., *Anal. Chem.,* **2015**, *87*, 1, 64-82

[32] Saito, K., Watanabe, Y., Shirakawa, M., Matsushita, Y., Imai, T., Koike, T., Sano, Y., Funada, R., Fukazawa, K., Fukushima, K., *Plant J.,* **2012**, *69*, 3, 542-552

[33] Hemalatha, R. G., Pradeep, T., *J. Agric. Food Chem.,* **2013**, *61*, 31, 7477-7487

[34] Skraskova, K., Heeren, R. M., *J. Chromatogr. A,* **2013**, *1319*, 1-13

[35] Sturtevant, D., Lee, Y. J., Chapman, K. D., *Curr. Opin. Biotechnol.,* **2016**, *37*, 53-60

[36] Lee, Y. J., Perdian, D. C., Song, Z., Yeung, E. S., Nikolau, B. J., *Plant J.,* **2012**, *70*, 1, 81-95

[37] Sarsby, J., Towers, M. W., Stain, C., Cramer, R., Koroleva, O. A., *Phytochemistry,* **2012**, *77*, 110-118

[38] Cha, S., Zhang, H., Ilarslan, H. I., Wurtele, E. S., Brachova, L., Nikolau, B. J., Yeung, E. S., *Plant J.,* **2008**, *55*, 2, 348-360

[39] Yoshimura, Y., Zaima, N., Moriyama, T., Kawamura, Y., *PLoS One,* **2012**, *7*, 2, e31285

[40] Fujii, T., Matsuda, S., Tejedor, M. L., Esaki, T., Sakane, I., Mizuno, H., Tsuyama, N., Masujima, T., *Nat. Protoc.,* **2015**, *10*, 9, 1445-1456

[41] Zenobi, R., *Science,* **2013**, *342*, 6163, 1243259

[42] Seaman, C., Flinders, B., Eijkel, G., Heeren, R. M., Bricklebank, N., Clench, M. R., *Anal. Chem.,* **2014**, *86*, 20, 10071-10077

[43] Fernie, A. R., Aharoni, A., Willmitzer, L., Stitt, M., Tohge, T., Kopka, J., Carroll, A. J., Saito, K., Fraser, P. D., DeLuca, V., *Plant Cell,* **2011**, *23*, 7, 2477-2482

[44] Erban, A., Schauer, N., Fernie, A. R., Kopka, J., in *Metabolomics: methods and protocols.* (Ed.: Weckwerth, W.), Humana Press, Totowa, New Jersey, **2007**, pp. 19–38.

[45] Kim, H. K., Verpoorte, R., *Phytochem. Anal,* **2010**, *21*, 1, 4-13

[46] Vuckovic, D., *Anal. Bioanal. Chem.,* **2012**, *403*, 6, 1523-1548

[47] Ernst, M., Silva, D. B., Silva, R. R., Vencio, R. Z., Lopes, N. P., *Nat. Prod. Rep.,* **2014**, *31*, 6, 784-806

[48] Glauser, G., Grata, E., Dubugnon, L., Rudaz, S., Farmer, E. E., Wolfender, J. L., *J. Biol. Chem.,* **2008**, *283*, 24, 16400-16407

[49] Shroff, R., Vergara, F., Muck, A., Svatos, A., Gershenzon, J., *Proc. Natl. Acad. Sci. U.S.A.,* **2008**, *105*, 16, 6196-6201

[50] De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., Hall, R. D., *Nat. Protoc.,* **2007**, *2*, 4, 778-791

[51] Oikawa, A., Otsuka, T., Jikumaru, Y., Yamaguchi, S., Matsuda, F., Nakabayashi, R., Takashina, T., Isuzugawa, K., Saito, K., Shiratake, K., *J. Sep. Sci.,* **2011**, *34*, 24, 3561-3567

[52] Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H.-M., Hesse, H., Segu, S., Hummel, J., Kohl, K., Willmitzer, L., *Plant J.,* **2011**, *68*, 2, 364-376

[53] Yanes, O., Tautenhahn, R., Patti, G. J., Siuzdak, G., *Anal. Chem.,* **2011**, *83*, 6, 2152-2161

[54] Badri, D. V., Chaparro, J. M., Zhang, R., Shen, Q., Vivanco, J. M., *J. Biol. Chem.,* **2013**, *288*, 7, 4502-4512

[55] Zhou, B., Xiao, J. F., Tuli, L., Ressom, H. W., *Mol. Biosyst.,* **2012**, *8*, 2, 470

[56] Boccard, J., Veuthey, J. L., Rudaz, S., *J. Sep. Sci.,* **2010**, *33*, 3, 290-304

[57] Khakimov, B., Bak, S., Engelsen, S. B., *J. Cereal Sci.,* **2014**, *59*, 3, 393-418

[58] Holcapek, M., Jirasko, R., Lisa, M., *J. Chromatogr. A,* **2012**, *1259*, 3-15

[59] Vaclavik, L., Mishra, A., Mishra, K. B., Hajslova, J., *Anal. Bioanal. Chem.,* **2013**, *405*, 8, 2671-2683

[60] Fuhrer, T., Zamboni, N., *Curr. Opin. Biotechnol.,* **2015**, *31*, 73-78

[61] Figueiredo, A., Fortes, A. M., Ferreira, S., Sebastiana, M., Choi, Y. H., Sousa, L., Acioli-Santos, B., Pessoa, F., Verpoorte, R., Pais, M. S., *J. Exp. Bot.,* **2008**, *59*, 12, 3371-3381

[62] Mahrous, E. A., Farag, M. A., *J. Adv. Res.,* **2015**, *6*, 1, 3-15

[63] Kim, H. K., Choi, Y. H., Verpoorte, R., *Trends in Biotechnol,* **2011**, *29*, 6, 267-275

[64] Bohni, N., Queiroz, E. F., Wolfender, J.-L., in *Encyclopedia of Analytical Chemistry* (Ed.: Meyers, R. A.), Wiley, Chichester, UK, **2014**, pp. 1–31.

[65] Sumner, L. W., Mendes, P., Dixon, R. A., *Phytochemistry,* **2003**, *62*, 6, 817-836

[66] Nakabayashi, R., Saito, K., *Anal. Bioanal. Chem.,* **2013**, *405*, 15, 5005-5011

[67] Theodoridis, G. A., Gika, H. G., Want, E. J., Wilson, I. D., *Anal. Chim. Acta,* **2012**, *711*, 7-16

[68] Luedemann, A., von Malotky, L., Erban, A., Kopka, J., *Methods Mol. Biol.,* **2012**, *860*, 255-286

[69] Christians, U., Klawitter, J., Hornberger, A., Klawitter, J., *Curr. Pharm. Biotechnol.,* **2011**, *12*, 7, 1053-1066

[70] Luedemann, A., Strassburg, K., Erban, A., Kopka, J., *Bioinformatics,* **2008**, *24*, 5, 732-737

[71] Lange, E., Tautenhahn, R., Neumann, S., Gropl, C., *BMC Bioinformatics,* **2008**, *9*, 375

[72] Tautenhahn, R., Bottcher, C., Neumann, S., *BMC Bioinformatics,* **2008**, *9*, 504, 1-16

[73] Lommen, A., Kools, H. J., *Metabolomics,* **2012**, *8*, 4, 719-726

[74] Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., Siuzdak, G., *Anal. Chem.,* **2006**, *78*, 3, 779-787

[75] Katajamaa, M., Oresic, M., *J. Chromatogr. A,* **2007**, *1158*, 1-2, 318-328

[76] Prince, J. T., Marcotte, E. M., *Anal. Chem.,* **2006**, *78*, 17, 6140-6152

[77] Sugimoto, M., Kawakami, M., Robert, M., Soga, T., Tomita, M., *Current Bioinformatics,* **2012**, *7*, 96-108

[78] Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J. J., Yanes, O., *Metabolites,* **2012**, *2*, 4, 775-795

[79] van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., van der Werf, M. J., *BMC Genomics,* **2006**, *7*, 142

[80] Worley, B., Powers, R., *Current Metabolomics,* **2013**, *1*, 1 92-107

[81] Mak, T. D., Laiakis, E. C., Goudarzi, M., Fornace, A. J. J., *Anal. Chem.,* **2014**, *86*, 1, 506-513

[82] Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. A., Hendriks, M. M. W. B., *Metabolomics,* **2013**, *10*, 3

[83] Alonso, A., Marsal, S., Julia, A., *Front. Bioeng. Biotechnol.,* **2015**, *3*, 23

[84] Shaffer, J. P., *Annu. Rev. Psychol.,* **1995**, *46*, 561-584

[85] Benjamini, Y., Hochberg, Y., *J. R. Stat. Soc. Series B Stat. Methodol.,* **1995**, *57*, 1, 289-300

[86] Gentleman, R. C*., et al.*, *Genome Biol.,* **2004**, *5*, 10, R80

[87] R Development Core Team, **2008**, http://www.r-project.org, accessed 14.01.2014

[88] Tautenhahn, R., Patti, G. J., Rinehart, D., Siuzdak, G., *Anal. Chem.,* **2012**, *84*, 5035-5039

[89] Tautenhahn, R., Cho, K., Uritboonthai, W., Zhu, Z., Patti, G. J., Siuzdak, G., *Nat. Biotechnol.,* **2012**, *30*, 9, 826-828

[90] Wishart, D. S*., et al.*, *Nucleic Acids Res.,* **2013**, *41*, Database issue, 801-807

[91] Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., *Nucleic Acids Res.,* **2014**, *42*, Database issue, D199-205

[92] Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., Merrill, A. H. J., Murphy, R. C., Raetz, C. R. H., Russell, D. W., Subramaniam, S., *Nucleic Acids Res.,* **2007**, *35*, Database issue, D527-532

[93] Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P., Karp, P. D., *Nucleic Acids Res.,* **2012**, *40*, Database issue, D742-753

[94] Ihaka, R., Gentleman, R., *J. Comp. Graph. Stat,* **1996**, *5*, 3, 299-314

[95] Vinaixa, M., EMBO Practical course on metabolomics bioinformatics for life scientists, **2014**

[96] Kovàts, E. S., *Helv. Chim. Acta,* **1958**, *41*, 206, 1915-1932

[97] Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmuller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A. R., Steinhauser, D., *Bioinformatics,* **2005**, *21*, 8, 1635-1638

[98] Fukushima, A., Kusano, M., *Front. Plant Sci.,* **2013**, *4*, 73

[99] Fiehn, O., Robertson, D., Griffin, J., Werf, M., Nikolau, B., Morrison, N., Sumner, L. W., Goodacre, R., Hardy, N. W., Taylor, C., Fostel, J., Kristal, B., Kaddurah-Daouk, R., Mendes, P., Ommen, B., Lindon, J. C., Sansone, S.-A., *Metabolomics,* **2007**, *3*, 3, 175-178

[100] Sumner, L. W*., et al.*, *Metabolomics,* **2007**, *3*, 3, 211-221

[101] Krueger, S., Steinhauser, D., Willmitzer, L., Giavalisco, P., *Plant J.,* **2012**, *70*, 1, 39-50

[102] Kind, T., Fiehn, O., *BMC Bioinformatics,* **2007**, *8*, 105, 1-20

[103] SciFinder CAS, https://scifinder.cas.org, accessed 01.08.2016

[104] Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcantara, R., Darsow, M., Guedj, M., Ashburner, M., *Nucleic Acids Res.,* **2008**, *36*, Database issue, D344-350

[105] Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P., Steinbeck, C., *Nucleic Acids Res.,* **2016**, *44*, D1, D1214-1219

[106] PubChem Compound Database, http://pubchem.ncbi.nlm.nih.gov, accessed 04.03.2016

[107] Dictionary of Natural Products DNP 24.2, http://dnp.chemnetbase.com, accessed 08.01.2016

[108] Afendi, F. M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L. K., Saito, K., Kanaya, S., *Plant Cell Physiol.,* **2011**, *53*, 2, e1(1–12)

[109] Bueschl, C., Krska, R., Kluger, B., Schuhmacher, R., *Anal. Bioanal. Chem.,* **2013**, *405*, 1, 27-33

[110] Bueschl, C., Kluger, B., Lemmens, M., Adam, G., Wiesenberger, G., Maschietto, V., Marocco, A., Strauss, J., Bodi, S., Thallinger, G. G., Krska, R., Schuhmacher, R., *Metabolomics,* **2014**, *10*, 754-769

[111] Horai, H.*, et al.*, *J. Mass Spectrom.,* **2010**, *45*, 7, 703-714

[112] Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., Akiyama, K., Sakurai, T., Matsuda, F., Aoki, T., Hirai, M. Y., Saito, K., *Phytochemistry,* **2012**, *82*, 38-45

[113] Smith, C. A., O'Maille, G., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., Custodio, D. E., Abagyan, R., Siuzdak, G., *Ther. Drug Monit.,* **2005**, *27*, 6, 747-751

[114] Schmelzer, K., Fahy, E., Subramaniam, S., Dennis, E. A., *Methods Enzymol.,* **2007**, *432*, 171-183

[115] Fahy, E., Sud, M., Cotter, D., Subramaniam, S., *Nucleic Acids Res.,* **2007**, *35*, Web Server issue, W606-612

[116] Kind, T., Liu, K. H., Lee do, Y., DeFelice, B., Meissen, J. K., Fiehn, O., *Nat. Methods,* **2013**, *10*, 8, 755-758

[117] Gerlich, M., Neumann, S., *J. Mass Spectrom.,* **2013**, *48*, 3, 291-298

[118] Morreel, K., Saeys, Y., Dima, O., Lu, F., Van de Peer, Y., Vanholme, R., Ralph, J., Vanholme, B., Boerjan, W., *Plant Cell,* **2014**

[119] Silva, R. R., Jourdan, F., Salvanha, D. M., Letisse, F., Jamin, E. L., Guidetti-Gonzalez, S., Labate, C. A., Vencio, R. Z. N., *Bioinformatics,* **2014**, *30*, 9, 1336-1337

[120] Stanstrup, J., Gerlich, M., Dragsted, L. O., Neumann, S., *Anal. Bioanal. Chem.,* **2013**, *405*, 15, 5037-5048

[121] Plant Metabolic Network (PMN), http://www.plantcyc.org, accessed 04.05.2014

[122] Chagoyen, M., Pazos, F., *Brief Bioinform,* **2013**, *14*, 6, 737-744

[123] Baart, G. J. E., Martens, D. E., *Methods Mol. Biol.,* **2012**, *799*, 107-126

[124] Cottret, L., Wildridge, D., Vinson, F., Barrett, M. P., Charles, H., Sagot, M.-F., Jourdan, F., *Nucleic Acids Res.,* **2010**, *38*, Web Server issue, W132-137

[125] Itkin, M., *et al.*, *Science,* **2013**, *341*, 6142, 175-179

— Chapter 2 —

Metabolomic characterization of the disease resistance gene *Lr34*

Metabolomic characterization of the disease resistance gene *Lr34*

1 Introduction

The world's population is estimated to grow to 9 – 10 billion by the year 2050 [1]. The concomitantly increased food demand requires an increase in crop-, particularly cereal production [2]. Maize, rice and wheat are amongst the most important cereal crops in the world (Figure 1)[3]. To sustainably ameliorate crop production and to limit harvest losses, the tolerance of plants for abiotic (e.g. drought, heat, cold) and biotic stress (e.g. weeds, insects, bacteria, fungi) must be improved. Amongst a variety of strategies to prevent yield loss, a promising approach is the use of disease resistance in crops.

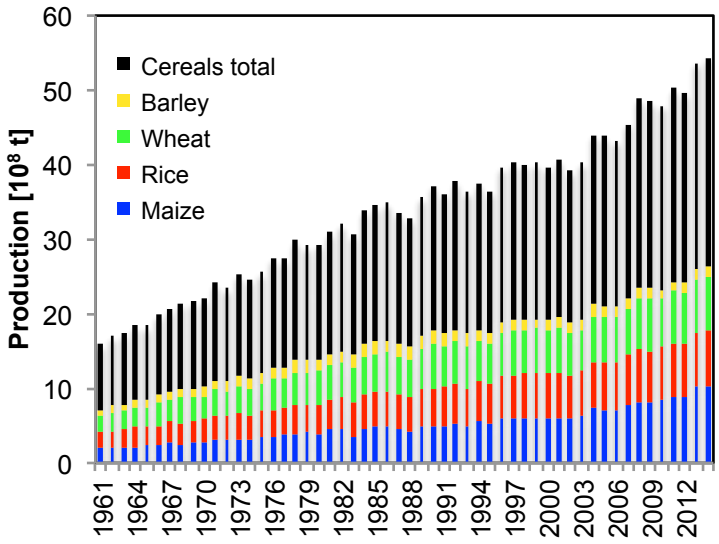


Figure 1: World production of cereal crops 1961–2014 from Food and Agriculture Organization of the United Nations [3].

1.1 Plant disease caused by fungal pathogens

Every year, substantial losses of crop yields are caused by fungal pathogens, and rusts are among the most damaging diseases of wheat [4], because of their rapid development, their ability to form new races, and their wide and long-distance distribution [5]. For instance, up to 40% of yield losses was reported in wheat as a consequence of leaf rust disease [6].

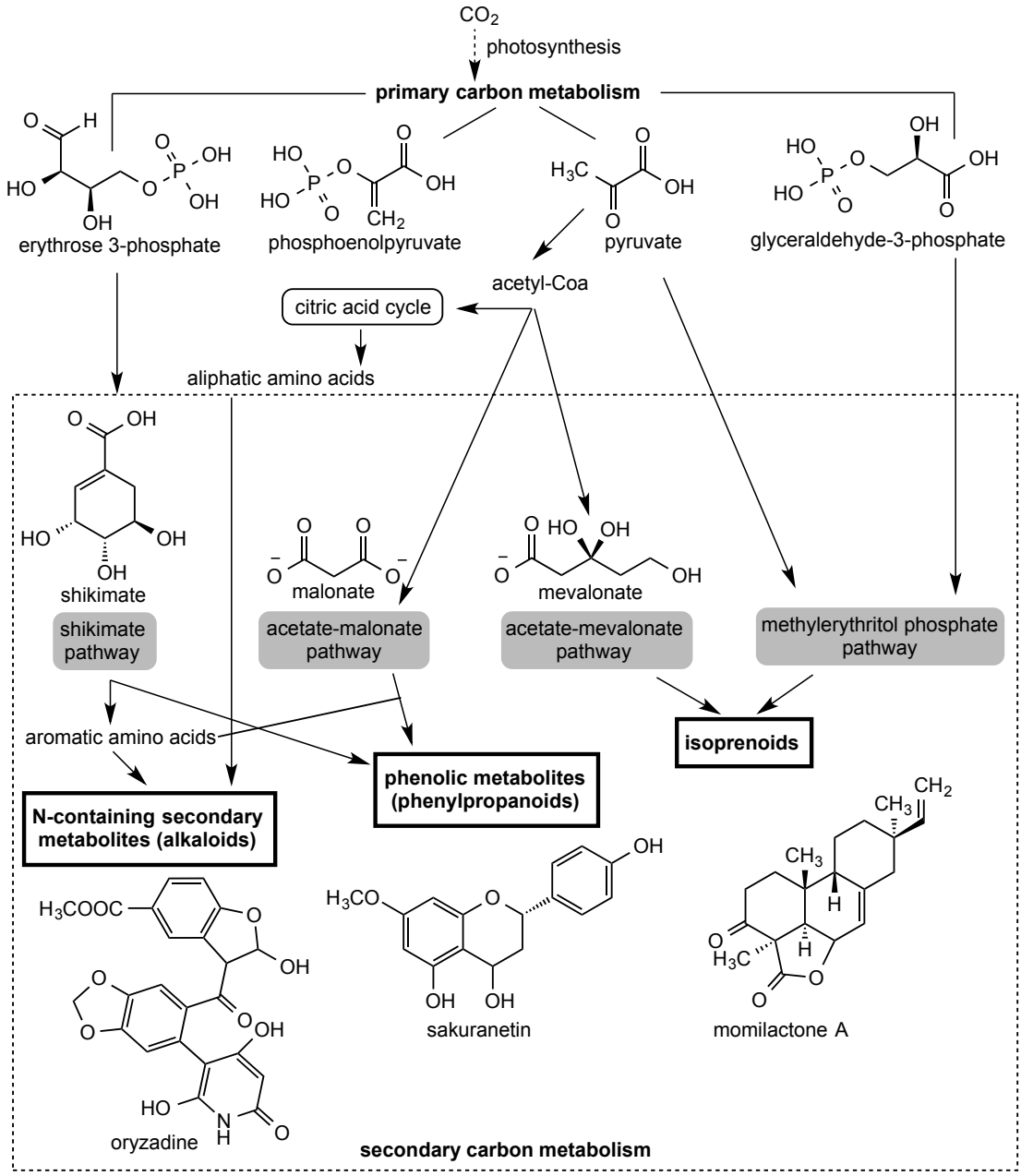
The fungal wheat diseases leaf-, stripe- and stem rust are caused by the pathogens *Puccinia triticina*, *Puccinia striiformis* f. sp. *tritici*, *Puccinia graminis* f. sp. *tritici*, and powdery mildew results from *Blumeria graminis* f.sp. *tritici* infection. Variants of stem and

stripe rust as well as powdery mildew also infect barley plants [7]. In rice, the most destructive fungal infection is rice blast disease, caused by *Magnaporthe oryzae* [7]. *Puccinia* as well as *Blumeria graminis* fungi develop an obligate biotrophic lifestyle, while *Magnaporthe oryzae* is hemi-biotrophic. Biotrophic fungi absolutely depend on a living host plant [8], and they form specialized structures situated inside plant cells to gain nutrients from living plant tissue [7]. In contrast, hemi-biotrophic fungi grow with a biotrophic lifestyle at the beginning of the infection of the plant, but after lesion development they switch to a more necrotrophic lifestyle that allows growing on dead plant tissue [9].

1.2 Disease resistance

Plants and pathogens interact in a two-way communication described as “zigzag” model [10], in which plants actively defend against pathogen attack with a two-tiered innate immune system [11]. In a first step, plant plasma membrane receptors recognize pathogen- or microbe associated molecular patterns (PAMP/MAMP) and initiate a general plant defense response, referred to as PAMP-triggered immunity (PTI). PAMPS are for example complex carbohydrates from fungal cell walls, proteins like the bacterial flagellin, or lipopolysaccharides present at outer cell membrane of Gram-negative bacteria. In turn, pathogens suppress PTI components by inserting so-called effector proteins directly to the cytoplasm, which suppress the host plant defense or benefit pathogen colonization [12]. In a second tier, plants sense specific pathogen effectors by resistance genes (R-genes) and trigger stronger secondary resistance response, called effector-triggered immunity (ETI). A specific feature of ETI is the activation of hypersensitive response, leading to localized cell death in order to limit pathogen spreading. Additionally, the accumulation of reactive oxygen species directly defends pathogens and activates signaling cascades, which in turn are based on phosphorylation cascades to modulate transcriptional regulation and phytohormone activation for further plant defense purposes [13].

Pathogens induce specific modulations of the plant’s primary and secondary metabolism, and they have been found to be highly specific for given tissues, species, and plant/pathogen interactions [14].



Scheme 1: Main biosynthetic pathways involved in the formation of defensive secondary metabolite classes (bold frame), adopted from [13, 14].

Defensive secondary plant metabolites are either constitutively present (phytoanticipins) or induced upon pathogen attack (phytoalexins). The three major groups of defensive secondary metabolites are alkaloids, isoprenoids and shikimates (Scheme 1) [14].

Based on this general interplay of plants and pathogens, different strategies for controlling infective diseases are used.

Control of fungal diseases is generally achieved by the deployment of fungicides and by the use of disease-resistant plant varieties. Host plant resistance is the preferred method

to protect cultivation and economic profit around the world [4], because resistance genes are environmental friendly, efficient and cost-effective [5].

Disease resistance can be attributed to three factors: genetic control, race specificity and durability. The genetic control of a disease resistance can either depend on a single (mono-), on several (oligo-), or on many genes (polygenic). While race specific resistance only works against some races of a particular pathogen, race non-specific resistance generally defeats all races.

Resistance is considered durable if despite a wide exposure of the plant to the pathogen, no increase of virulence of the pathogen is observed for considerable time period.

Two main categories of disease resistance are distinguished. Complete resistance to a pathogen race caused by a single gene (qualitative resistance) and incomplete resistance based on multiple genes (quantitative resistance). The efficacy of single R-genes is limited by the potential to be rapidly overcome by virulent pathogens [8]. As a prominent example, important resistance genes have been defeated by the very aggressive wheat stem rust race Ug99, named after its discovery in Uganda in 1999, which infects both barley and wheat and spread out from East Africa into Yemen and Iran [15]. The Ug99 race group was classified as a major threat to food security due to its fitness, virulence attributes and adaptive capacity [7].

Although many R-genes against rust are known, only a few have been cloned in barley and wheat [15].

For the combination of multiple genes with a partial effect, the resistance is usually more durable, because pathogens overcoming one of the genes involved only have a minor advantage [8]. Genes underlying quantitative disease resistance are referred to as quantitative resistance loci (QRL) [16]. The identification of partial resistance loci is a difficult task and requires expertise in breeding, large population sizes and specific combination of alleles [15]. But still, two such partial resistance loci have been cloned: *Lr34* [17] and *Yr36* [18] for leaf- and yellow (stripe) rust respectively. *Yr36* is encoding a kinase with a putative START lipid-binding domain and confers resistance to stripe rust at relatively high temperatures [18]. Another example for race nonspecific resistance is the recessive barley powdery mildew resistance gene *Mlo* [19], encoding an integral membrane protein.

So far, only a small number of nonspecific wheat resistance genes are known to confer partial resistance to multiple fungal species, including *Lr34*, *Lr46* [20] and *Lr67* [21].

1.3 Resistance gene *Lr34*

The wheat resistance gene *Lr34* (synonymous with the inferred pleiotropic effect genes *Yr18*, *Sr57*, *Pm38*) is the most important source of disease resistance against leaf- and stripe rust in wheat breeding due to its durability with respect to pathogen virulence. *Lr34* causes broad-spectrum resistance against the three wheat rusts and powdery mildew. Whereas pathogens overcome other resistance genes by rapid adaptation, *Lr34* is durable over several decades [22]. Although *Lr34*-mediated resistance is only partial, it reduces the frequency of infections, results in smaller colonies, a longer latency period and specifically reduces the intercellular hyphal growth of the obligate biotrophic pathogens [23].

In adult wheat plants, *Lr34* is expressed in the uppermost flag leaf [17], with highest transcript levels in the leaf tip, during the critical grain-filling stage.

Although extensively studied, little is known about the molecular mechanism of the *Lr34*-mediated resistance [24]. Keller and co-workers found that *Lr34* encode for a full-size ATP-binding cassette (ABC) transporter protein, LR34 [17]. The substrate(s) translocated by LR34 and also its location remain unknown.

Lr34 functionality can be transferred to other cereal species like barley, rice and sorghum; *Lr34* increased resistance against the species-specific diseases barley leaf rust, barley powdery mildew [25], rice blast [26] and sorghum rust (unpublished data). The resistance spectrum of *Lr34* thus covers biotrophic and hemi-biotrophic fungal pathogens infecting different cereal species. In contrast to wheat, where *Lr34* is only expressed at adult plant stage, rice [26], barley [27] and sorghum already showed *Lr34* resistance at seedling stage. Therefore, transgenic barley, rice and sorghum plants offer a valuable tool to study the molecular differences in pathogen resistance. The gene activity is correlated to leaf-tip necrosis [28], which serves as phenotypic marker for *Lr34*. Remarkably, this senescence-like process affecting leaf tips of adult wheat plants [29] does not reduce the overall wheat yield [30]. In contrast, *Lr34* causes severe a leaf-tip necrosis phenotype in barley, leading to poor plant growth and reduced seed production [25]. In rice, a transgenic line with late leaf-tip necrosis, high disease resistance level and no growth penalty was described [26].

Different alleles of *Lr34* originating from different wheat lines were compared [17, 22]. Of the 1401 amino acids of the corresponding complete proteins, only two differences were found between the resistant and the susceptible alleles. The first amino acid difference is a deletion of a phenylalanine probably in a trans-membrane region; the second difference is

a change from tyrosine to histidin predicted at the end of a trans-membrane helix [22]. The phenylalanine deletion was found to be sufficient to introduce the *Lr34*-based responses in barley [27].

Biochemical, transcriptional and cytological studies of *Lr34* wheat provided evidence that the resistance is reminiscent of some components of hypersensitive response [23], a plant defense reaction that restricts biotrophic and other pathogens [31], but without the accompanying localized cell death [24].

Although *Lr34* confers resistance towards fungal pathogens, *Lr34* containing plants were not responsive to pathogen inoculation [32], indicating constitutive rather than induced response.

1.4 ABC transporters

The protein LR34 was found to be a full-size ABC transporter of the ABCG subfamily (formerly called pleiotropic drug resistance transporter) [17]. ABC transporters are a superfamily of integral transmembrane proteins that act as adenosine triphosphate-driven pumps to translocate a set of diverse substrates across membranes, independent of concentration gradients and membrane potentials. In plant genomes, over one hundred different ABC transporters are encoded [33]. ABCG transporters are abundant in plants, but none had been identified in animals [32]. In both *Oryza sativa* (rice) and *Arabidopsis*, the G subfamily was the largest group of ABC transporters found [34].

In *Arabidopsis*, most known ABCG transporters are localized in the plasma membrane [33]. This subfamily had been reported to be involved in pathogen resistance (e.g. AtABCG36 in *Arabidopsis*), transport of signaling molecules (e.g. AtABCG25 and AtABCG40 in *Arabidopsis* for abscisic acid), volatile compounds, antimicrobial terpenoids, auxinic herbicides [34] and heavy metal resistance (e.g. AtABCG40), but also in exudate excretion or cuticle formation [32].

1.5 Metabolomics of cereals

Metabolomics found a variety of applications in plant science such as phenotyping of different species, analysis of resistance traits, responses to herbicides [35] or the equivalence of genetically modified and conventional crops [36]. Combined with transcriptomics, metabolomics is extensively used in crop breeding [37], e.g. to assess variances of natural varieties or cultivars. Many studies investigate metabolomic plant

stress response to a wide range of abiotic and biotic stressors. Overviews of recent metabolomics studies in the cereals wheat, barley and rice are listed in Table 1 – Table 3. Thereby, the devastating *Fusarium* head blight disease is one of the most studied subjects in barley and wheat [38]. Having a necrotrophic lifestyle, the *Fusarium* fungus excretes trichothecene mycotoxins to destroy and macerate the host plant tissue to feed on [39]. In rice, plant-pathogen interaction with rice blast was investigated [40, 41].

Upon pathogen infection of a plant, both partners alter their molecular and physiological status compared to the non-infected state. Plant-pathogen interactions can be distinguished on transcript- and protein level, but no specificity is exhibited on metabolome level [42]. Hence, the observed metabolic profile from a pathogen-infected plant is a mixture of the metabolomes of the interacting host and pathogen, sometimes referred to as dual metabolomics. Strategies to distinguish the contribution of both interaction partners involve isotopic labeling or in situ imaging of metabolites to achieve spatial resolution. As specific fungal biomarkers, the steroid ergosterol or arachidoic acid [43] were used in the past.
Considering the whole plant tissue, only a minor number of cells are infected with pathogen, thus representing considerable technical challenges [42].

Table 1: Metabolomics studies in wheat, with study factors including fungal pathogens highlighted in grey.

Study factors	Organ	Method					Detector	Metabo-lomics		Type of metabolomics	Ref
		LC-UV	GC-MS	LC-MS	HR	MS/MS		targeted	non-targeted		
Fusarium head blight infection	Rachis, spikelet		X				EI-TOF-MS		X	Primary metabolite profiling	[44]
Fusarium head blight infection	Rachis, spikelet			X	X		ESI-LIT-Orbitrap	X	X	Metabolite profiling	[45]
Fusarium head blight infection - decipher QTL (Fhb1) mechanism	Rachis, spikelet			X	X		ESI-LIT-Orbitrap	X	X	Metabolite profiling, proteomics	[46]
Phloem exudate, aphid herbivory	Phloem exudate		X	X			EI-Q-MS, ESI-QqQ-MS			Metabolite profiling	[47]
45 Wheat lines	Seed			X	X		ESI-Q-TOF		X	Metabolite profiling	[48]
Phenolic compound profiling of several wheat varieties	Grain			X	X		ESI-TOF-MS	X		Targeted metabolite profiling	[49]
Transgenic, antifungal genes (RIP, KP4)	Leaf	X		X		X	UV, APCI, ESI-IT-MS	X		Targeted flavonoid profiling	[50]
Transgenic wheat (glutenin Dx5 subunit)	Caryopses		X				EI-Q-MS		X	Metabolite profiling	[51]
Identify QTLs under water limitation	Leaf (flag)		X				EI-Q-MS	X		Primary metabolite profiling	[52]
Heat stress durum wheat	Seed		X				EI-Q-MS	X		Primary metabolite profiling	[53]
Drought stress	Leaf (flag)		X				EI-Q-MS	X		Primary metabolite profiling	[54]
Cytoplasmic genome substitution	Leaf		X				EI-Q-MS	X		Primary metabolite profiling	[55]
Genotype, environment and genotype-by-environment interaction	Grain		X				EI-Q-MS	X		Primary metabolite profiling	[56]
Benzoxazinoid derivatives	Grain			X	X	X	ESI-Q-TOF-MS	X		Benzoxazinoid derviatives	[57]

Table 2: Metabolomics studies in barley, with study factors including fungal pathogens highlighted in grey.

Study factors	Organ	Method					Detector	Metabo-lomics		Type of metabolomics	Ref
		LC-UV	GC-MS	LC-MS	HR	MS/MS		targeted	non-targeted		
Powdery mildew infection	Leaf			X			ESI-Q-MS	X		Primary metabolite profiling	[58]
Infection powdery mildew, root pathogen *Sebacinales*	Leaf	X		X			ESI-QqQ-MS	X		Targeted amino acid, carbohydrate, organic acids and	[59]
Powdery mildew infection	Leaf			X		X	ESI-ion trap-MS		X	Targeted secondary metabolites	[60]
Fusarium head blight infection	Spikelet			X	X	X	ESI-LIT-Orbitrap-MS		X	Metabolite profiling	[61]
Fusarium head blight infection	Spikelet			X	X	X	ESI-LIT-Orbitrap-MS, MS/MS		X	Metabolite profiling	[62]
Fusarium head blight infection	Spikelet			X	X	X	ESI-LIT-Orbitrap-MS		X	Metabolite profiling	[63]
Fusarium head blight infection	unclear			X	X	X	ESI-Q-TOF-MS	X	X	Fingerprinting, mycotoxin	[64]
Fusarium head blight infection	Spikelet			X	X	X	ESI-LIT-Orbitrap-MS		X	Metabolite profiling	[65]
Fusarium head blight infection	Spikelet			X	X	X	ESI-LIT-Orbitrap		X	Metabolite profiling	[66]
Fungi on barley, oats and rye	Grain		X				EI-TOF-MS, EI-FID	X		Mycotoxins, VOC, FAMES	[67]
Tunisian races	Grain	X					UV	X		Targeted total phenolics, flavonoids, tannins	[68]
Transgenic moss Na^+ transporter	Leaf		X				EI-Q-MS	X		Primary metabolite profiling	[69]
Transgenic barley (*β*-1,3-Glu, cThEn)	Leaf	X		X			UV, ESI-QLIT-MS	X	X	Targeted and untargeted metabolite profiling, transcript	[70]
Drought stress	Root	X		X			ESI-Q-MS, UV	X		Targeted metabolites	[71]
Vacuolar transport	Leaf	X					UV	X		Targeted flavonoid analysis	[72]
Phenolics in barley- and flaxseed	Hull			X	X	X	ESI-Q-TOF-MS		X	Profiling of major phenolics	[73]
Natural varieties	Leaf			X	X	X	ESI-IT-MS^n, Orbitrap-MS/MS, NMR	X		Phenolic secondary metabolites	[74]
Storage metabolome vacuole	Protoplast		X	X	X	X	EI-TOF-MS, ESI-FT-ICR-MS		X	Metabolite profiling	[75]
Nitrogen deficiency	Shoot, root			X		X	ESI-QqQ-MS	X		Targeted analysis soluble phenols, amino acids	[76]
Phosphate deficiency	Shoot, root		X	X			EI-Q-MS, ESI-IT-MS	X		Primary metabolite profiling	[77]
Salt stress	Leaf, root		X				EI-Q-MS	X		Primary metabolite profiling	[78]
Salt stress	Leaf, root		X				EI-Q-MS	X		Primary metabolites	[79]
Boron toxicity	Leaf, root		X				EI-QqQ-MS		X	Primary metabolite profiling	[80]
Comparison wheat, barley, oat, rye	Grain		X				EI-Q-MS		X	Primary metabolite profiling	[81]

Table 3: Metabolomics studies in rice, with study factors including fungal pathogens highlighted in grey.

Study factors	Organ	Method					Detector	Metabo-lomics		Type of metabolomics	Ref
		LC-UV	GC-MS	LC-MS	HR	MS/MS		targeted	non-targeted		
Rice blast infection	Leaf		X	X		X	EI-Q-MS, ESI-Q-TOF-MS, NMR-500 MHz		X	Metabolite profiling	[41]
Rice blast infection	Leaf			X			EI-TOF-MS, DI-LIT-MS	X	X	Metabolite fingerprinting + profiling	[40]
Bacterial leaf blight infection	Leaf		X	X	X		EI- and ESI-TOF-MS		X	Metabolite profiling	[82]
Response to herbivory	Leaf			X	X	X	ESI–TOF–MS, ESI-QqQ-MS/MS	X	X	Metabolite profiling, targeted phenolamides and momilactones analysis	[83]
Comparison of 175 rice accessions	Aerial part			X	X	X	ESI-Q-TOF-MS		X	Secondary metabolite profiling	[84]
Tissue specific and developmentally controlled flavonoid accumulation	Leaf, various			X	X	X	ESI–LIT –MS/MS	X		Widely targeted metabolomics	[85]
Metabolic shifts during grain development	Seed		X	X	X	X	EI-Q-MS, ESI-Q-TOF-MS		X	Metabolite profiling	[86]
Metabolic variation between cultivars	Seed		X	X	X	X	ESI-LIT -MS, ESI-LIT-FTICR-MS, EI-Q-MS		X	Metabolite profiling	[87]
Network nitrate limitation	Leaf, root		X	X			ESI-LIT-MS, EI-Q-MS		X	Metabolite profiling, transcripts	[88]
Different pigmented rice seeds, antioxidant activity	Seed		X	X	X	X	ESI-Q-TOF-MS, ESI-IT-MS/MS, EI-TOF-MS		X	Metabolite profiling	[89]
Suspension cell culture as model for salinity tolerance	Suspension cultured cells		X				EI-Q-MS		X	Metabolite profiling, proteomics	[90]
Translational metabolomics of Arabidopsis, rice, tomato and Chlamy	Leaf, root			X		X	ESI-IPC–QqQ-MS/MS	X		Ion suppression and recovery tests	[91]
Chromium stress	Root		X				EI-Q-MS, NMR-400 MHz	X		Primary metabolite profiling, transcripts	[92]
Transgenic (Bt, CpTI)	Grain		X				EI-Q-MS	X		Primary metabolite profiling	[93]
Comparison of cultivars	Seed		X				EI-TOF-MS, EI-GCxGC-TOF-MS	X		Primary metabolite profiling	[94]
Transgen Bt, β-1,3-Glu	Grain	X	X				UV, EI-Q-MS	X		Primary metabolites, protein, elements	[95]
Fe transporter knockdown mutant	Leaf, root		X				EI-TOF-MS	X		Primary metabolites	[96]
Salt stress	Leaf, root		X				EI-Q-MS	X		Primary metabolites	[97]
Drought markers in rice cultivars	Leaf		X		X		EI-TOF-MS		X	Primary metabolite profiling	[98]
Polyamine metabolism, drought	Leaf	X	X		X		EI-TOF-MS, fluorescence	X	X	Primary metabolite profiling, polyamine analysis	[99]

2 Aim

In order to gain a deeper insight into the metabolic pathways and responses related to *Lr34*, metabolomic characterization of *Lr34* was performed as part of a multi-disciplinary approach to unravel the molecular function of durable disease resistance in cereals. In previous studies [22-25, 100], the *Lr34* gene has been studied on the level of gene expression. The mayor aim of the work presented here was to characterize the complex interplay of a multitude of metabolites involved in *Lr34* pathogen resistance utilizing metabolomics based approaches.

The first aim was to investigate whether *Lr34* caused differences in the metabolomic phenotype by comparing metabolic profiles or fingerprints of control plants with *Lr34* containing plants. A further goal was to recognize characteristic differences between control plants and *Lr34* plants and to identify the responsible metabolites in order to converge the biological relevance of *Lr34*-mediated resistance. Finally, the purpose was to embed the identified metabolites into pathways to correlate the metabolic response of *Lr34* in context of defense response, and to compare the *Lr34* responses of different cereal species and growth conditions.

The current chapter of metabolic characterization of *Lr34* is structured into two parts, of which the first part was investigated transgenic *Lr34* barley (3, Metabolic fingerprinting of transgenic *Lr34* barley by LC-MS). The major aim of this first series of experiments was to focus on secondary metabolites and to compare control plants with *Lr34* containing barley plants, and thus the metabolic fingerprinting was restricted to UHPLC-MS based technology.

For the second part, a more extensive series of experiments aimed at expanding the metabolic characterization of *Lr34* in terms of cereal species, growth conditions and additional classes of metabolites (4, Metabolomic profiling of *Lr34* resistance in barley, rice and wheat by LC- and GC-MS). Therefore, not only transgenic *Lr34* barley, but also transgenic *Lr34* rice and natural *Lr34* wheat were investigated. Three different growth conditions were examined for barley and rice including hydroponic plant growth on nutrient solution, non-infected control growth conditions and plants infected with fungal pathogens. In order to account for robustness of plant growth, two experiments of each condition were

performed. To circumvent the effects of the senescence phenotype, barley and rice genotypes with low and high expression levels of *Lr34* were investigated. Furthermore, *Lr34* wheat grown under most natural field conditions was evaluated. To cover a broad variety of plant metabolites including primary metabolites, lipids, secondary metabolites and targeted phytohormones, a combination of UHPLC- and GC-MS based technologies were applied to investigate metabolic profiles and fingerprints.

3 Metabolic fingerprinting of transgenic *Lr34* barley by LC-MS

To gain knowledge about the metabolic response of *Lr34*-mediated broad-spectrum resistance, a number of experiments were performed following the general metabolomics workflow described in chapter 1.
In wheat, *Lr34* confers resistance only in adult plants and is expressed at highest levels in the flag leaf [23], thereby requiring a full growth season to produce suitable plant material. As an alternative, transgenic *Lr34* barley was available, that displays a clear, early phenotype occuring along with rust resistance at very young seedling stage [27]. Thus, transgenic barley offered the possibility to gain plant material for metabolic fingerprinting within a month's time frame, and it further allowed growth under standardized hydroponic conditions.

For our metabolic fingerprinting investigation, we compared a group of transgenic *Lr34* barley plants to a group of control plants from wild type or sister lines that contain no transgene. Generally, the plants were grown under controlled conditions, whole leaves of seedlings were harvested, flash frozen in liquid nitrogen to quench metabolism, and homogenized into plant powder. For the extraction of the plant powder, we used an extraction protocol which is routinely used for LC-MS metabolomic profiling of plants [101], with special focus on semipolar secondary metabolites such as alkaloids, saponins, phenolic acids, phenylpropanoids, flavonoids, glucosinolates, polyamines and derivatives thereof. The extracts were analyzed by ultra-high-performance liquid chromatography coupled to mass spectrometry (UHPLC-MS). UHPLC-MS data were preprocessed by XCMS to generate a two-dimensional matrix including features (characterized by *m/z* - RT combination) and the corresponding intensities of the individual replicate samples. For each of the experiments, different statistical methods were used. In all experiments, statistically significant features differing between the metabolic fingerprints of transgenic *Lr34* barley and controls were observed. Subsets of these significantly different features were selected for metabolite identification. The identified metabolites could act as possible relevant metabolic biomarkers to describe the metabolic response of *Lr34* in barley and might potentially be candidates for the unknown transported substrate of the LR34 transporter.

Five metabolic fingerprinting experiments were performed with transgenic *Lr34* barley by LC-MS, and a summary of the experimental condition is listed in Table 4. The table describes the plant material used, including the transgenic and control barley lines, the growth environment and medium, the harvesting time point and the homogenization procedure. Furthermore, details about the extraction procedure and LC-MS analysis are summarized.

Experiment Exp1 was performed with barley plants grown in the greenhouse that were harvested after 21 days, and with an extraction procedure that was marginally different from the other experiments. Experiment Exp2 – Exp4 used similar procedures for the preparation of the plant material, but extraction and chromatography of Exp2 was slightly different. While plants from experiment Exp1 – Exp4 were grown on hydroponic nutrient solution to control the growth environment including the growth medium, soil as more natural growth medium was selected for experiment Exp5 to verify changes of metabolites under natural growth conditions. Additionally, two transgenic *Lr34* lines – BG8 and BG9, which differ in the *Lr34* expression level – were selected for Exp5.

All experiments Exp1 – Exp4 showed significant differences based on *XCMS Online* processing (Supplementary table 1). To investigate whether control and transgenic barley could be separated based on the multivariate analysis of data from metabolic fingerprinting, principal component analysis (PCA) was performed for each experiment and each ionization mode (Supplementary figure 1 and Supplementary figure 2). A clear group separation was achieved in Exp1 and Exp3, which indicates the potential discriminating power of the statistically significant ions. For experiments Exp2 and Exp4, no clear separation between control and transgenic *Lr34* groups was obtained. Therefore, the differences in metabolic fingerprints between controls and *Lr34* of experiments Exp1 and Exp3 were further investigated and secondary metabolites were identified.

Table 4: Experimental parameters for plant growth, sample preparation and metabolite analysis for metabolic profiling of transgenic *Lr34* barley by LC-MS.

Experiment	Plant material					Extraction				LC-MS	
	Barley lines*	Growth environment	Growth medium	Harvest	Homo-genization	Solvent	Ratio	Procedure	Dilution	LC column dimension	MS
Exp1	Wt vs BG8	greenhouse	hydroponic	day 21, 3rd leaf	grindmill, 45s, 30Hz	MeOH : H_2O 8:2 + 0.1% HCOOH	1:10	grindmill, 5min, 30Hz, filter units	-	1 x 50mm	(–)- and (+)-ESI
Exp2	Sib vs BG8	phytotron	hydroponic standard procedure	day 14, 3rd leaf	grindmill, 2x 25s, 25Hz	MeOH + 0.1% HCOOH	1:10	vortex 20s, filter units	-	1 x 50mm	(–)- and (+)-ESI
Exp3	Sib vs BG8	phytotron	hydroponic standard procedure	day 14, 3rd leaf	mortar + pestle	MeOH + 0.1% HCOOH	1:3	sonicate 15min 20°C, cfg	1:1 H_2O	2.1 x 100mm	(–)- and (+)-ESI
Exp4	Sib vs BG9	phytotron	hydroponic standard procedure	day 16, 3rd leaf	mortar + pestle	MeOH + 0.1% HCOOH	1:3	sonicate 15min 20°C, cfg	1:1 H_2O	2.1 x 100mm	(–)- and (+)-ESI, hardware upgrade
Exp5	Sib8, Sib9 vs BG8, BG9	phytotron	soil	day 18, 3rd leaf 2nd leaf (BG8, BG9)	mortar + pestle	MeOH + 0.1% HCOOH	1:3	sonicate 15min 20°C, cfg	1:1 H_2O	2.1 x 100mm	(–)- and (+)-ESI

3.1 Putative annotation of upregulated cyanoglycoside derivatives in Exp3

To determine significant metabolic features, characterized by a combination of *m/z* and RT, different statistical analyses were performed based on *XCMS Online*, *MetaboLyzer* and *Feature filtering* of both (–)- and (+)-ESI data (Table 5).

Table 5: Significant features detected in Exp3 with different statistical processing methods

ESI	Statistical processing method	Total	Significance / Intensity		Fold change	
		No. features	Criterium	No. features	Criterium	No. features
(–)	*XCMS Online*	879	Welch's *t*–test, unequal variances, *p*-value ≤ 0.05	205	FC > 1.5	121
	MetaboLyzer	953	Welch's *t*-test, *p*-value ≤ 0.05	382	FC > 1.5	344
	Feature filtering	233	CV(samples) < CV(QC), intensity > MS/MS threshold, Student's *t*-test, *p*-value ≤ 0.05	57	FC >1.5	3
(+)	*XCMS Online*	2393	Welch's *t*–test, unequal variances, *p*-value ≤ 0.05	365	FC > 1.5	240
	MetaboLyzer	2346	Welch's *t*-test, *p*-value ≤ 0.05	575	FC > 1.5	103
	Feature filtering	996	CV(samples) < CV(QC), intensity > MS/MS threshold, Student's *t*-test, *p*-value ≤ 0.05	106	FC >1.5	32

For a single metabolic component, several adducts or fragments corresponding to individual features may occur. Thus, the number of detected features only gives a limited indication about the number of regulated metabolites.

Generally, a higher number of significant features were detected by processing with *XCMS Online* and *MetaboLyzer*, while a lower number of features was obtained by *Feature filtering*. The latter applies more restrictive selection criteria and only highlights the most intense and robust changes in the dataset. Therefore, the three features found in LC-(–)-ESI-MS data processing by *Feature filtering* were selected as a starting point for metabolite annotation and completed with a selection of upregulated components detected with the different statistical analyses (Table 6). The table summarizes a variety of features that cluster into components **a**–**f** according their retention time, their similarity in significance and fold change, and their similar MS/MS fragmentation behavior. Within the clusters, the *m/z* of the features could be correlated with known mass shifts of commonly observed ESI-adducts, and the components were later annotated as metabolites. The detection of the selected components as various adducts in both (+)- and (–)-ESI ionization modes demonstrates the consistency and importance of the underlying signals.

The significance of the statistical analysis was listed as *p*-value, additionally to the fold change and the intensity of the individual features.

Table 6: Selection of significantly upregulated components in Exp3. Components **a**–**c** and **f** were obtained from statistical processing, while * components **d**, **e** and **g** were observed in extracted ion chromatograms. The *p*-value and fold change (FC) are based on *XCMS Online*, the intensity values are means of sister line; detection of features with statistical processing was done by *XCMS Online* (1), *MetaboLyzer* (2) and *Feature filtering* (3).

Component		Feature							Statistical processing		
#	Formula	Ion	*m/z*	RT [min]	Δ *m/z* [ppm]	*p*-value	FC	Intensity	1	2	3
a	$C_{11}H_{17}NO_7$	$[M+HCOO]^-$	320.0982	1.10	-1.5	6.9E-03	2.3	8995	X	X	X
		$[M+H]^+$	276.1077	1.12	-0.2	2.1E-02	2.2	5998	X	X	X
		$[M+NH_4]^+$	293.1342	1.12	-0.3	4.4E-04	2.0	2856	X		
		$[M+Na]^+$	298.0896	1.12	-0.5	5.3E-02	1.7	3107	X		
b	$C_{11}H_{19}NO_6$	$[M+HCOO]^-$	306.1189	1.85	-1.8	7.1E-03	2.4	15169	X	X	X
		$[M+Hiso]^+$	263.1310	1.87	-3.2	5.8E-02	2.0	638	X	X	X
		$[M+NH_4]^+$	279.1549	1.87	-0.4	1.2E-03	2.3	4753	X		X
		$[M+Na]^+$	284.1103	1.87	-0.6	3.5E-02	2.1	3164	X		X
c	$C_{11}H_{17}NO_6$	$[M+HCOO]^-$	304.1034	2.21	-1.4	1.0E-02	2.6	3495	X	X	X
		$[M-H]^-$	258.0967	2.21	-6.5	1.2E-02	1.8	328	X	X	
		$[M+H]^+$	260.1127	2.23	-0.6	2.8E-02	2.7	2304	X		X
		$[M+NH_4]^+$	277.1393	2.24	-0.3	5.5E-03	2.7	835	X		
		$[M+Na]^+$	282.0946	2.23	-0.7	3.3E-02	2.3	2112	X		
d*	$C_{11}H_{19}NO_6$	$[M+HCOO]^-$	306.1190	2.42	-1.3	-	-	784			
e*	$C_{11}H_{17}NO_6$	$[M+HCOO]^-$	304.1034	3.40	-1.3	1.6E-02	2.4	678	X		
f	$C_{11}H_{19}NO_6$	$[M-H]^-$	260.1134	3.91	-1.9	7.0E-03	2.7	1131	X	X	
		$[M+HCOO]^-$	306.1190	3.91	-1.4	6.7E-03	2.5	12180	X	X	
		$[M+H]^+$	262.1282	3.92	-1.2	4.3E-02	2.3	530	X		
		$[M+NH_4]^+$	279.1549	3.92	-0.5	1.4E-04	2.1	3789	X		
		$[M+Na]^+$	284.1102	3.92	-0.8	3.6E-02	2.0	2368	X		X
g*	$C_{11}H_{19}NO_6$	$[M+HCOO]^-$	306.1193	4.14	-0.3	-	-	1484			

Extracted ion chromatograms of the significant features detected in (–)-ESI-MS by *Feature filtering*, $[M+HCOO]^-$ of components **a**, **b** and **c**, are shown in Figure 2. In the EIC of *m/z* 306.119, another intense peak of component **f** was observed which was found to significantly increase in transgenic *Lr34* barley plants. Additionally, two low abundant peaks **d** and **g** were recognized in the EIC of component **b**, but not detected by any of the statistical methods. In the EIC of *m/z* 304.104, a low abundant peak of component **e** was

found, which was only detected to significantly change by *XCMS Online*. The additionally observed features **d** and **g** were characterized by MS/MS fragmentation, but not statistically analyzed due to low abundance.

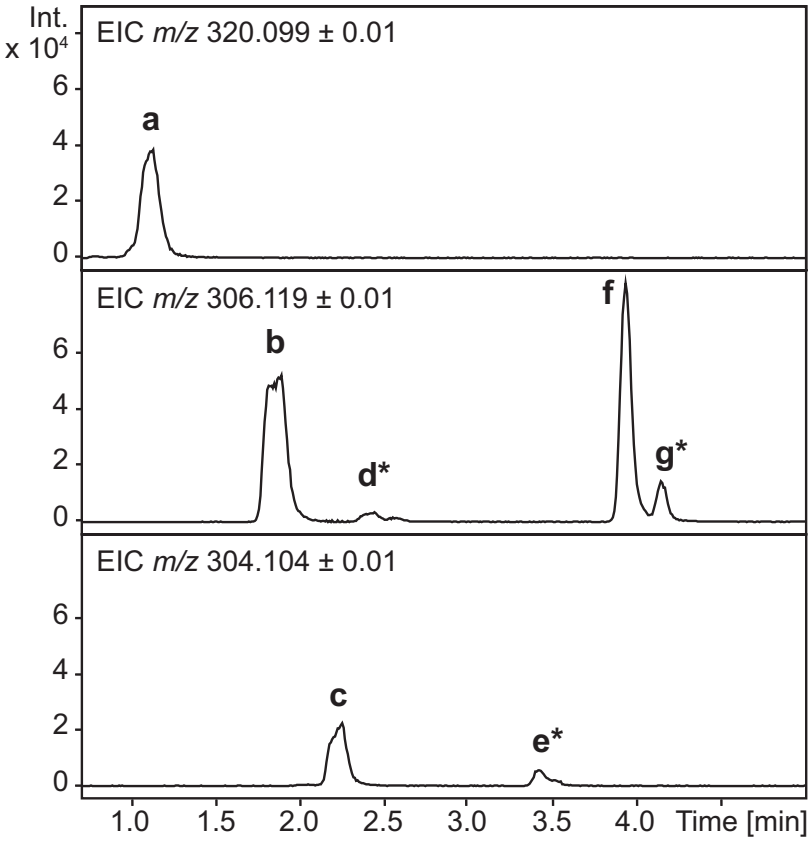


Figure 2: (–)-EIC of formiate adducts of components **a–g** in experiment Exp3, pool of BG8. Components **a**–**c** and **f** were obtained from statistical processing, while * components **d**, **e** and **g** were observed in extracted ion chromatograms.

Fold-change (FC) is an important parameter to characterize the degree of metabolite change in perturbed biological systems [102]. It is the ratio between the mean peak areas of the sister line (Sib) and the transgenic *Lr34* barley line (BG8). With fold changes of around two, a clear upregulation can be recognized for components **a**–**f** in the transgenic line. Both LC-(+)-ESI-MS and LC-(–)-ESI-MS acquisition methods showed similar fold changes and statistical significance based on Welch's *t*-test, that compared means of two independent sample groups assuming that the variances of the groups differ [103]. Looking at the extracted ion chromatograms of the most intense feature, $[M+HCOO]^-$ of component **b**, a significant upregulation was observed for *Lr34* compared to the control despite variations between individual replicates of Sib and BG8 (Figure 3).

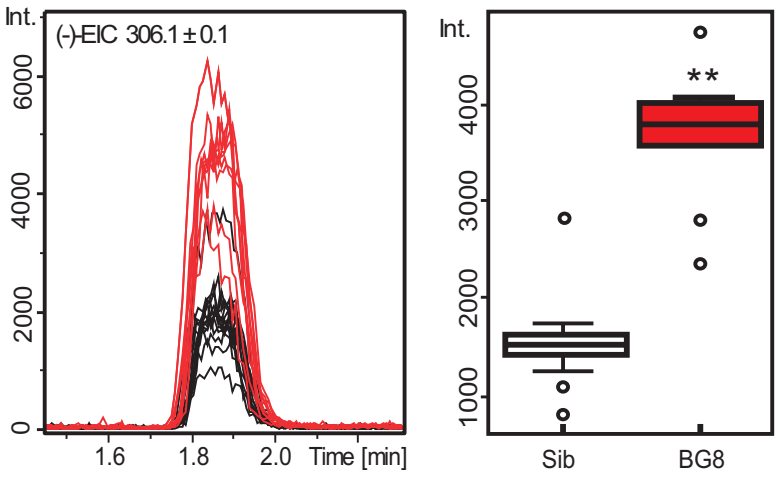


Figure 3: Significantly upregulation of feature $[M+HCOO]^-$ of component **b** (*m/z* 306.1 at RT 1.85 min) with signals of Sib (black) and transgenic BG8 (red) found in experiment Exp3: (–)-EIC left, *XCMS Online* boxplot of peak areas right, *p*-value **< 0.01 based on Welch's *t*-test.

To evaluate the consistency between the experiments, the signal intensities of the protonated features of components **a**–**c** and **f** were inspected in all experiments Exp1–Exp4 (Figure 4). The most significant upregulation was observed in experiment Exp3, while less significant increases of the concentrations of components **a**–**c** and **f** were observed in *Lr34* plants of the other experiments Exp1, Exp2 and Exp4.

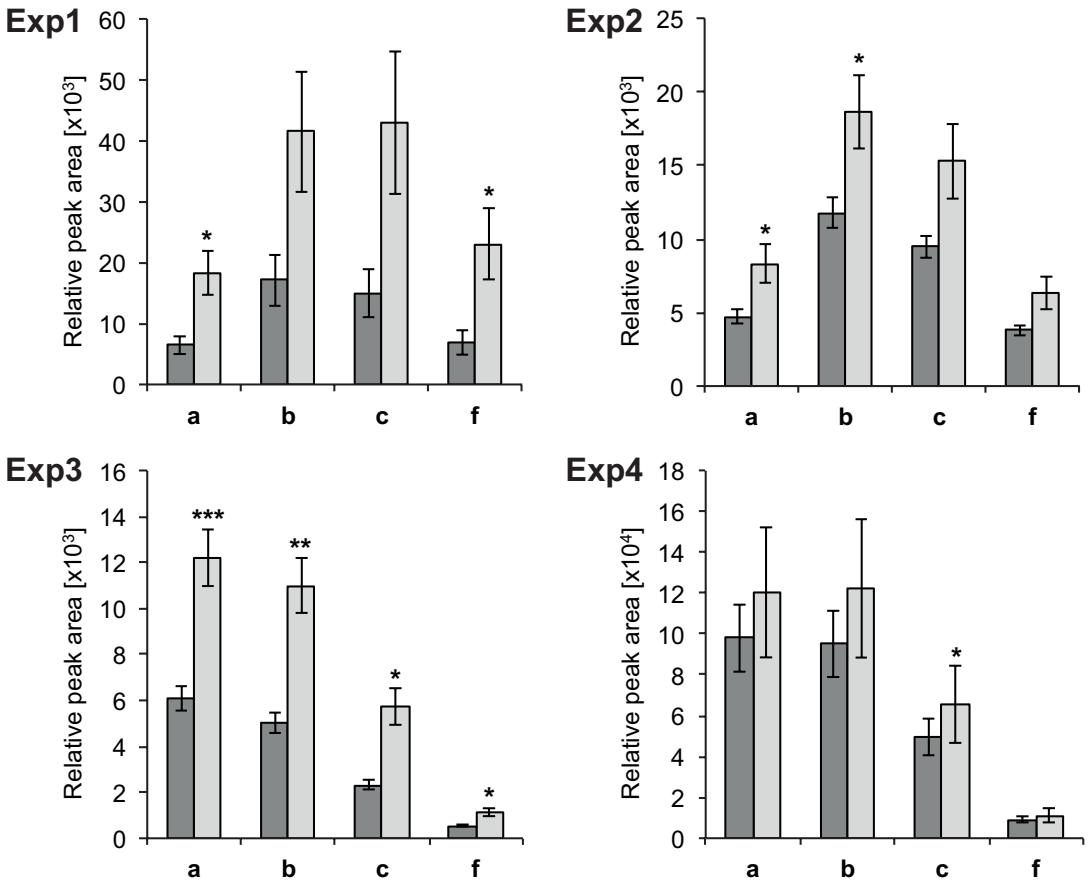


Figure 4: Relative quantification of components **a**–**c** and **f** in barley seedlings based on $[M+H]^+$ EICs in experiments Exp1–Exp4. Control (dark grey) and transgenic *Lr34* lines (light grey) signal intensities of peak areas are given relative to internal standard ampicillin, values are means and error bars represent standard error (* *p*-value <0.05, ** < 0.01 and *** < 0.001 based on Student's *t*-test).

In order to investigate barley plants grown in a natural environment on soil, Exp5 was designed. In contrast to previous experiments, two transgenic barley lines with different levels of *Lr34* expression were evaluated. In the third leaves, only BG8 showed a significant increase of components **a**–**c** and **f**, while highly significant differences were observed in the second leaves of both BG8 and BG9 lines (Figure 5).

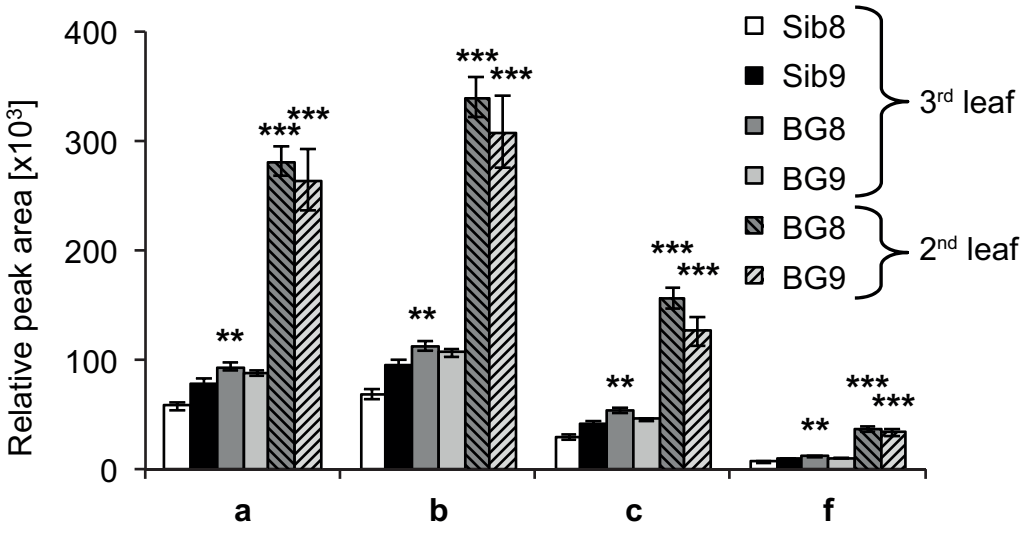


Figure 5: Relative quantification of components **a**–**c** and **f** in barley seedlings based on $[M+H]^+$ EICs in experiments Exp5. Peak areas are given relative to internal standard ampicillin, values are means and error bars represent standard error (* *p*-value <0.05, ** < 0.01 and *** < 0.001 based on Student's *t*-test).

After highlighting significantly different components between control and transgenic *Lr34* barley by statistical methods, the goal was to identify the responsible metabolites. By combining the HR-MS data from components **a**–**g** with additionally recorded CID MS/MS spectra of the formiate and ammonium adducts, molecular formulas and substructures of the metabolites were characterized. As an example, the problem of metabolite annotation is illustrated for component **b** in the following section.

Table 7: Calculation of molecular formulas of the $[M+HCOO]^-$ feature of component **b** (*m/z* 306.1191 at 1.85 min), based on *SmartFormula3D* algorithm with restrictions $C_aH_bN_cO_dPS$ $1 \leq b/a \leq 3$; e=0 or 1; f=0 or 1; a, b, c, and d not limited, rings and double bonds 0.5–40.

SmartFormula	*m/z* calculated	Mass error [mDa]	Mass error [ppm]	mSigma
$C_{12}H_{20}NO_8$	306.1194	0.34	1.10	2.8
$C_9H_{12}N_{11}O_2$	306.1181	-1.01	-3.31	7.8
$C_7H_{17}N_9O_3P$	306.1198	0.64	2.09	15.0

In a first step, CID MS/MS spectra of the formiate and ammonium adducts of component **b** were recorded in both ionization modes. In a second step, the calculation of the molecular formulas were performed based on HR-MS data of the formiate adduct of component **b**. Three molecular formulas for the ion at *m/z* 306.1198 were suggested by the

SmartFormula3D algorithm, which restricts the number of formulas by considering fragments as subsets of the respective molecular formulas (Table 7). The *mSigma* value characterizes the fit of the relative isotopic abundance, and values below 10 are considered reliable. Accordingly, the third formula $C_7H_{17}N_9O_3P$ with *mSigma* value of 15 was considered unlikely. The mass deviation of the second formula $C_9H_{12}N_{11}O_2$ with 3.31 ppm was above the deviations of internal standards (e.g. camphor sulfonic acid Δ -0.2 ppm). Additionally, the N/C-atom ratio of 1.2 was at the upper end of a range of N/C 0–1.3 suggested by the seven golden rules for heuristic filtering of molecular formula [104]. For the first formula $C_{12}H_{20}NO_8$, both a low mass error of 1.1 ppm and a low *mSigma* value of 2.8 were determined. Several features detected at RT 1.85 min, that were suggested to belong as adducts to the same component, supported the proposed chemical formula $C_{11}H_{19}NO_6$ of component **b** with low mass errors.

Analogous to component **b**, molecular formula were assigned to components **a** ($C_{11}H_{17}NO_7$), **c** and **e** ($C_{11}H_{17}NO_6$), and **d, f** and **g** ($C_{11}H_{19}NO_6$) which were isomers of component **b**.

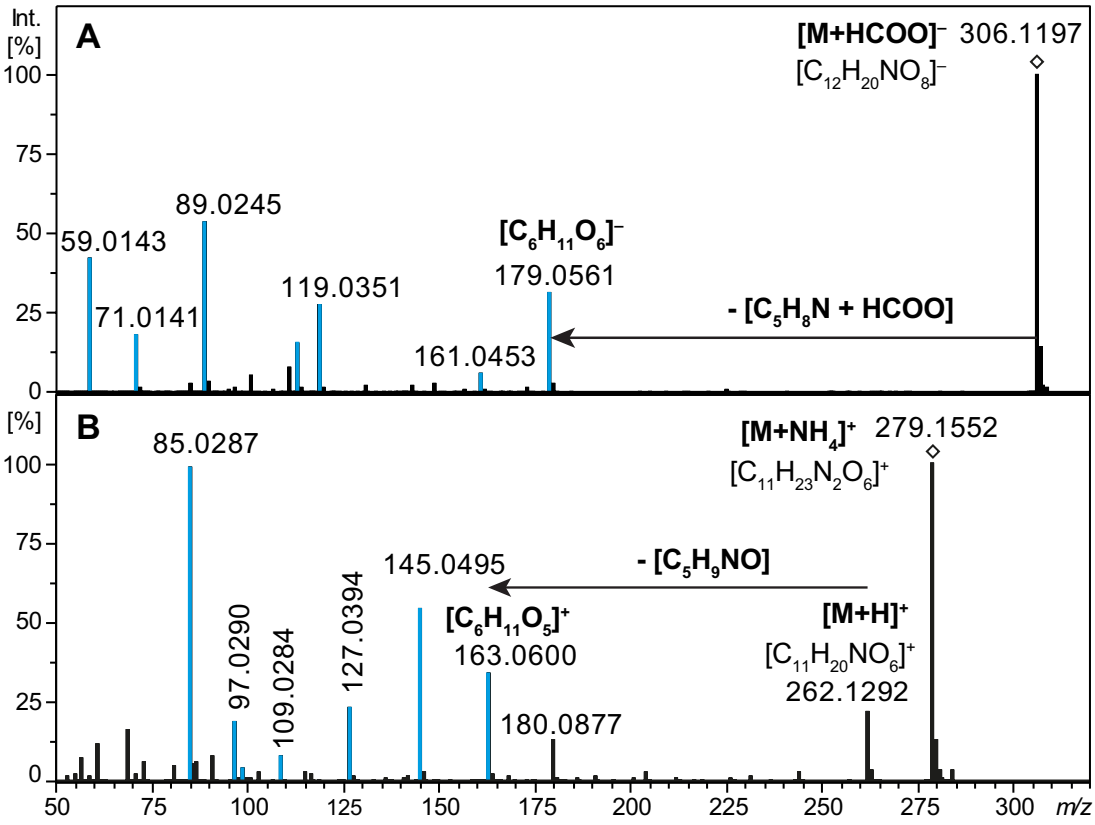


Figure 6: CID MS/MS of component **b**, (A) (–)-ESI MS/MS of *m/z* 306 (CE 10–35 eV) and (B) (+)-ESI MS/MS *m/z* 279 (CE 5–35 eV). Blue fragments were in agreement with *MassBank* spectra for sugar molecules.

The tandem mass spectra that were obtained for component **b** by (–)- and (+)-ESI-MS/MS (Figure 6) were matched with the spectral databases *MassBank* [105] and *ReSpect* [106]. Numerous MS/MS fragments with $m/z \leq 179$ in negative and $m/z \leq 163$ in positive ionization mode showed a high agreement with database spectra of the isobaric sugars glucose, fructose, mannose and galactose. Thus, the metabolite of interest is most probably a conjugate of one of these sugars. The mass difference of the supported sugar fragments $[C_6H_{11}O_6]^-$ and $[C_6H_{11}O_5]^+$ to the respective precursor ions suggested that the remaining part of the analyte represents a nitrogen-containing moiety of the chemical formula C_5H_8N. This portion of the molecule did not yield any own fragments in both ionization modes. So far, component **b** was presumed to be a metabolite containing a glycosyl unit and a yet unknown structural part with formula C_5H_8N.

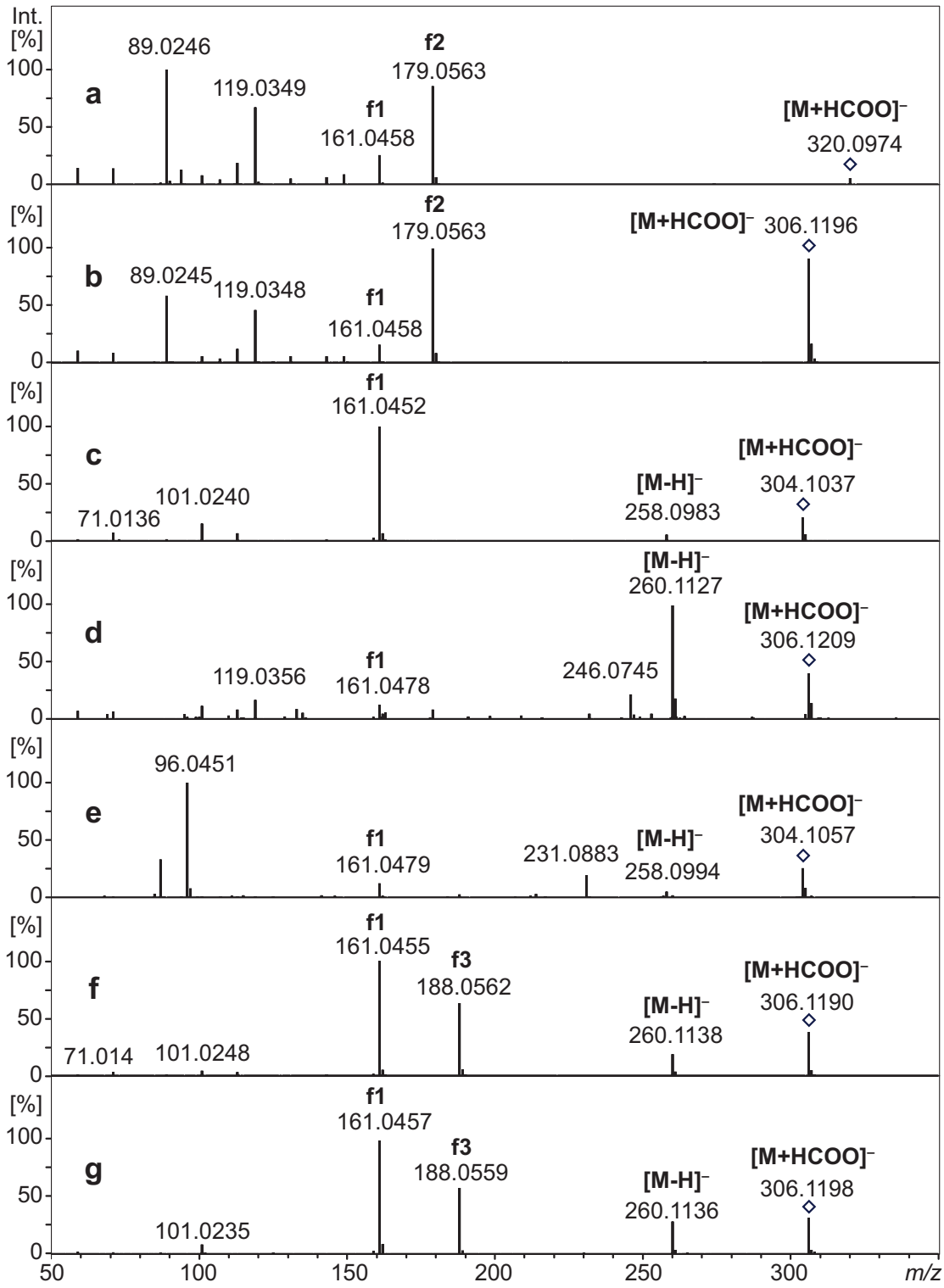


Figure 7: (–)-ESI CID-MS/MS of the formiate adducts of components **a**–**g** listed in Table 9 (CE 15–35 eV), fragment interpretations: f1 $[C_6H_9O_5]^-$, f2 $[C_6H_{11}O_6]^-$, f3 $[C_7H_{10}NO_5]^-$.

The same was true for components **a**, **c**-**g**, which also contained a glycosyl unit and mass differences corresponding to the formulas C_5H_6NO (**a**), C_5H_8N (**d**, **f**, **g**) or C_5H_6N (**c**, **e**) (Table 8).

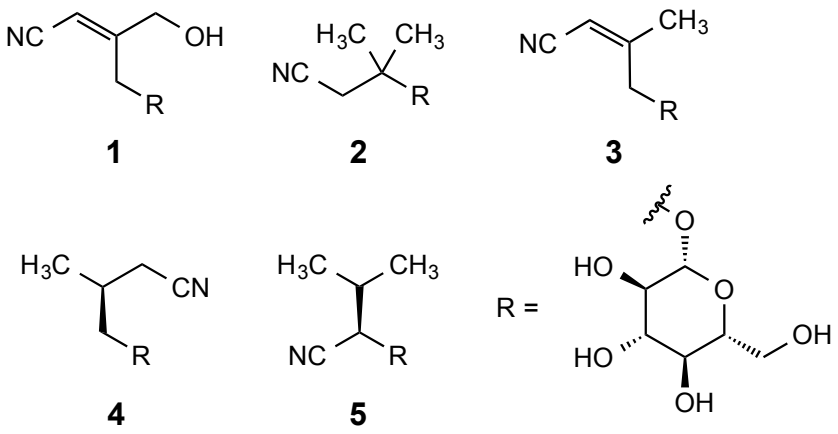
Table 8: Chemical formulas of components **a**–**g** in (–)-ESI glycosyl part and neutral losses

Component	Chemical formula			
	Metabolite	Formiate adduct	Glycosyl fragment	Neutral loss
a	$C_{11}H_{17}NO_7$	$[C_{12}H_{18}NO_9]^-$	f2 $[C_6H_{11}O_6]^-$	C_5H_6NO + HCOO
b	$C_{11}H_{19}NO_6$	$[C_{12}H_{20}NO_8]^-$	f2 $[C_6H_{11}O_6]^-$	C_5H_8N + HCOO
c	$C_{11}H_{17}NO_6$	$[C_{12}H_{18}NO_8]^-$	f1 $[C_6H_9O_5]^-$	$C_5H_6N + H_2O$+ HCOO
d	$C_{11}H_{19}NO_6$	$[C_{12}H_{20}NO_8]^-$	f1 $[C_6H_9O_5]^-$	$C_5H_8N + H_2O$ + HCOO
e	$C_{11}H_{17}NO_6$	$[C_{12}H_{18}NO_8]^-$	f1 $[C_6H_9O_5]^-$	$C_5H_6N + H_2O$ + HCOO
f	$C_{11}H_{19}NO_6$	$[C_{12}H_{20}NO_8]^-$	f1 $[C_6H_9O_5]^-$	$C_5H_8N + H_2O$ + HCOO
g	$C_{11}H_{19}NO_6$	$[C_{12}H_{20}NO_8]^-$	f1 $[C_6H_9O_5]^-$	$C_5H_8N + H_2O$ +HCOO

The commercially available linamarin and lotaustralin were used as reference compounds for the comparison of MS/MS spectra obtained for components **a**–**g**, although they were not reported in barley but in white clove and cassava [107, 108]. Fragmentations of the reference compounds linamarin and lotaustralin showed a fragment at *m/z* 188.0557 corresponding to $[C_7H_{10}NO_5]^-$, which was supposed to result form rearrangement of the cyano group to the glucosyl moiety, and was indicative for the cyano group in α-position to the glucose. The fragment f3 $[C_7H_{10}NO_5]^-$ was also observed for compounds **f** and **g**. No further structural information regarding the cyanoalkyl moiety was obtained due to lack of fragmentation.

Although the mass spectra did not allow the assignment of a specific glycosyl unit, all previously reported cyano glycosides in barley hinted towards glucose.

In order to gain further information about possible molecular structures of the non-glycosyl part, the MS/MS spectra were loaded into the *MetFusion* tool [109], which combines library search with *in silico* fragmentation (*MetFrag* [109]). Two cyano glycosides (heterodendrin, lotaustralin) and an amino acid derivative (mycosporine) were thus suggested as possible candidates. A database query done in *SciFinder* with the chemical formula $C_{11}H_{19}NO_6$ and the keyword “barley” revealed different structures of cyano glucosides. Cyano glycosides are well known to accumulate in barley seedlings [110]. The cyano glycosides sutherlandin (**1**), epidermin (**2**), osmaronin (**3**), dihydroosmaronin (**4**) and epiheterodendrin (**5**) found in barley are biosynthetically derived from the amino acid *L*-leucine (Scheme 2)[110]. The non-glycosidic parts of cyano glycosides **1**–**5** matched with the mass differences observed in (–)-ESI MS/MS of components **a**–**d** and **f**.



Scheme 2: Cyano glucosides reported in barley; sutherlandin (**1**), epidermin (**2**), osmaronin (**3**), dihydroosmaronin (**4**), epiheterodendrin (**5**)

On the basis of a literature reference [110], which used similar chromatographic separation conditions as applied in the current study, the relative retention times of the components **a**–**d** and **f** could be correlated with the cyano glycosides **1**–**5** by molecular mass and succession of elution (Table 9). Components **e** and **g** were observed by [110], but not further characterized.

Table 9: List of cyano glycosides with putative assignments.

Com-ponent	Putative assignment	Molecular formula	MW [g mol^{-1}]	RT [110] [min]	RT [min]	Reference in barley
a	Sutherlandin **1**	$C_{11}H_{17}NO_7$	275.3	3.0	1.12	[110-114]
b	Epidermin **2**	$C_{11}H_{19}NO_6$	261.3	4.0	1.85	[110] [111-114]
c	Osmaronin **3**	$C_{11}H_{17}NO_6$	259.3	6.5	2.24	[110-113]
d	Dihydro-osmaronin **4**	$C_{11}H_{19}NO_6$	261.3	8.0	2.42	[110-113]
e	unknown	$C_{11}H_{17}NO_6$	259.3	13.0	3.43	[110]
f	Epihetero-dendrin **5**	$C_{11}H_{19}NO_6$	261.3	13.5	3.95	[110-112, 114]
g	unknown	$C_{11}H_{19}NO_6$	261.3	14.0	4.14	[110]

We were able to correlate chromatographic information from literature with our own data and to putatively identify five cyano glycosides whose structure was described previously, and to further describe two cyano glycosides that were mentioned in literature but not characterized by MS/MS before.

For components **a**–**d** and **f**, we found reference data in literature (metabolite identification level 2, chapter 1, 2.6), while components **e** and **g** were assigned based on to analogies of MS/MS fragmentation (metabolite level 3).

To close the circle of the metabolomics workflow, the increase in cyano glycoside concentration in *Lr34* barley was correlated with information about the biological relevance of cyano glycosides.

Generally, the function of cyano glycosides in barley is not yet fully understood and thus the interpretation of the increased cyano glycoside concentrations in context with *Lr34* is difficult. Cyano glycosides are separated in two categories — cyanogenic and non-cyanogenic — according to their ability to cleave off hydrogen cyanide. Living organisms release hydrogen cyanide in response to biotic or abiotic stress, a phenomenon that is observed throughout the plant kingdom and provides plants with immediate chemical defense response [107]. Cyano glycosides were generally classified as phytoanticipins [115], which are constitutive defense compounds being constantly synthesized by the plant. Increased levels of cyano glycosides were observed at ambient and high temperatures after powdery mildew infection of barley seedlings [60]. The cyano glycoside compound class had been described in interaction with powdery mildew as defense compounds, recognition factors or scavengers of reactive oxygen species [60, 110]. Of the cyano glycosides found in barley, only epidermin is cyanogenic and can be cleaved by the respective enzymes, *β*-glucosidases. Surprisingly, *β*-glucosidases are harbored in many plants, but barley lacks such enzymes [112]. When a *β*-glucosidase from sorghum was expressed in barley, cyanogenesis was reconstituted [107].

One possible hypothesis is that barley is able to release hydrogen cyanide in another way than enzymatically. For example, this could be purely chemically based on acid catalysis, which was supported by the structural rearrangement observed under (–)-ESI MS/MS conditions for epiheterodendrin (component **f**), linamarin and lotaustralin. Alternatively, *Lr34* plants might utilize other mechanisms for protecting itself against biotrophic fungal pathogens, and cyano glycosides in barley might have gained new functions.

The main location of cyano glycosides in barley leaves was reported to be the outermost cell layer of the leaf, the epidermis [107], but homogeneous distribution throughout the barley leaf surface had been observed by imaging mass spectrometry [112]. Indeed, the epidermis is also the main cell layer colonized by the biotrophic barley powdery mildew

fungus, whereas the other three fungal pathogens leaf, stripe and stem rust colonialize mesophyll cells of the host plant.

The absence of cyanogenesis in barley suggests an alternative function of cyano glycosides than as phytoanticipins [107]. Cyano glycosides contain 90% of the soluble carbohydrates in barley epidermis [107], and might therefore be involved in metabolic turnover as sugar or nitrogen storage compounds, like for example investigated in the rubber tree [116]. Multiple functions of cyano glycosides have been reported, including nitrogen and carbohydrate storage and transport, roles in seed dormancy and germination and hypothetically in regulation of reactive oxygen species, which might influence plant defense response [116].

3.2 Putative annotation of upregulated hordatines in Exp1

Experiment Exp1 showed a clear separation of the control and the transgenic *Lr34* barley in PCA (Supplementary figure 1). After statistical processing, an eminent cluster of six components that significantly differ in the abundance in *Lr34* versus the control barley plants was detected in Exp1. The respective components, showing fold changes of up to 6.6, were detected at retention times around 2.0–2.3 min (Table 10).

Table 10: Selection of significantly upregulated components in Exp1; *p*-value, fold change (FC) based on *XCMS Online*, intensity based on mean of Wt; statistical processing with *XCMS Online* (1) and *Feature filtering* (2).

Component		Feature							Statistical processing	
No.	Formula	Ion	*m/z*	RT [min]	Δ *m/z* [ppm]	*p*-value	FC	Intensity	1	2
i	$C_{34}H_{48}N_8O_9$	$[M+2H]^{2+}$	357.1839	2.03	1.2	1.5E-02	3.3	28682	X	X
j	$C_{35}H_{50}N_8O_{10}$	$[M+2H]^{2+}$	372.1893	2.02	0.2	3.8E-02	2.6	55623	X	X
k	$C_{36}H_{52}N_8O_{11}$	$[M+2H]^{2+}$	387.1947	2.08	-1.4	4.5E-02	1.8	9730		X
l	$C_{28}H_{38}N_8O_4$	$[M+2H]^{2+}$	276.1577	2.35	1.4	7.1E-03	6.6	191187	X	X
		$[M-H]^-$	549.2926	2.38	-3.2	1.1E-01	5.4	1466		
m	$C_{29}H_{40}N_8O_5$	$[M+2H]^{2+}$	291.1629	2.32	0.8	7.6E-03	4.8	190111	X	X
		$[M-H]^-$	579.3030	2.37	-3.2	1.9E-01	3.2	1373		
n	$C_{30}H_{42}N_8O_6$	$[M+2H]^{2+}$	306.1681	2.37	1.5	1.8E-02	3.1	40073	X	X

The extracted ion chromatograms of components **i**–**n** are shown in Figure 8. For each component, double peaks were observed that were not distinguished by the peak picking algorithm and therefore not separately evaluated in statistical analysis.

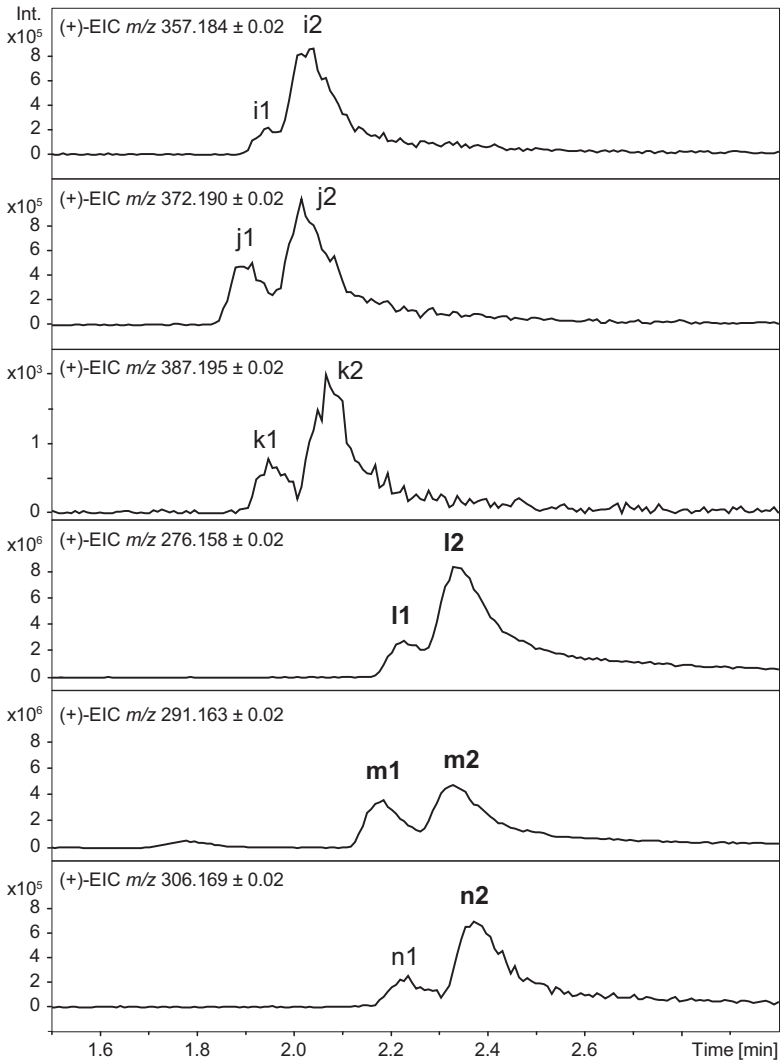


Figure 8: (+)-EIC of $[M+2H]^{2+}$ of components **i**–**n** in experiment Exp1, BG8. For the bold components, CID-MS/MS spectra of the $[M+H]^{+}$ ions are available.

In a next step, the metabolic differences found of components **i**–**n** in experiment Exp1 were investigated in all experiments Exp1–Exp4. The signal intensities of components **i**–**n** were evaluated based on the EICs of the $[M+2H]^{2+}$ ions, and the double peaks were summarized (Figure 8). The most pronounced significance and the highest increase of components **i**–**n** was observed in Exp1, whereas a tendency of increase between control and transgenic *Lr34* plants was observed in the subsequent experiments Exp2–Exp4.

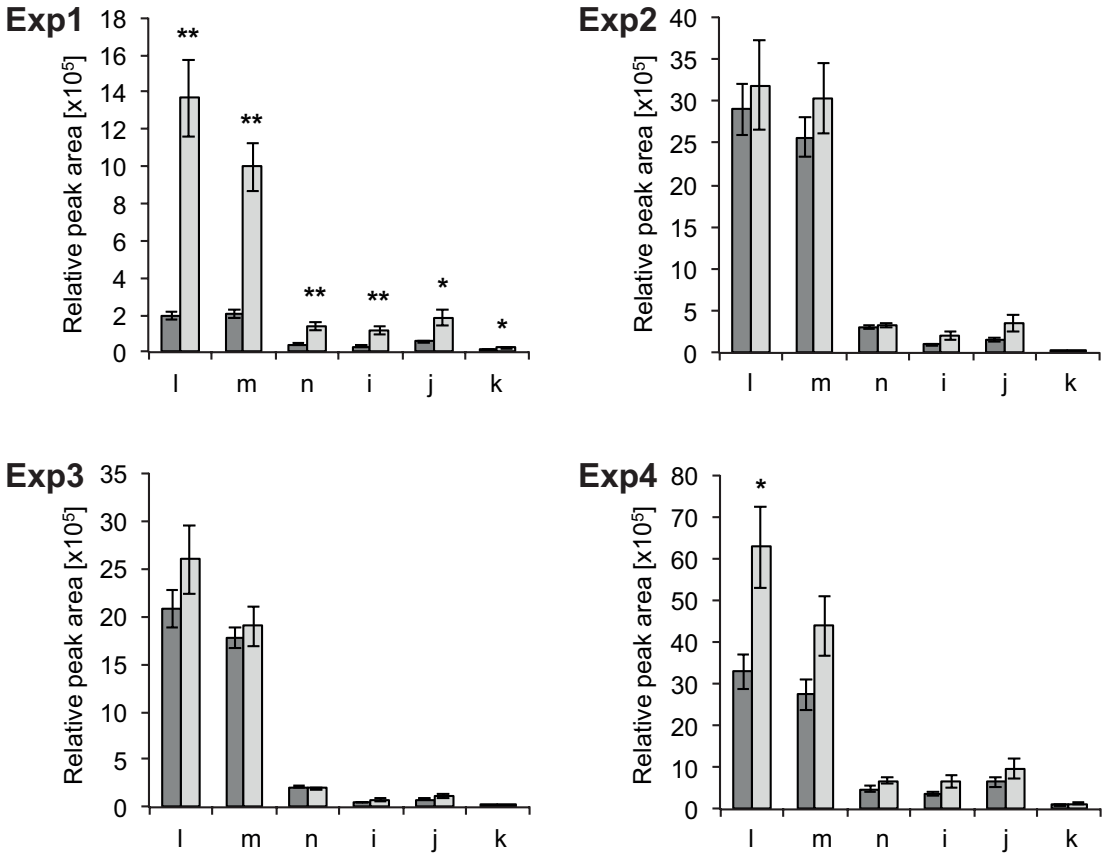


Figure 9: Relative quantification of components **i**–**n** in barley seedlings based on $[M+2H]^{2+}$ EICs in experiment Exp1–Exp4. Control (dark grey) and transgenic *Lr34* lines (light grey) signal intensities are given relative to internal standard ampicillin, values are means and error bars represent standard error (* *p*-value <0.05, ** < 0.01 and *** < 0.001 based on Student's *t*-test).

In order to verify the observed trend in a more natural environment, barley plants were grown on soil in experiment Exp5, in which two transgenic lines with different levels of *Lr34* expression were studied (Figure 10). Component **m** and **n** were significantly upregulated in third leaves of high *Lr34* expressing plants of line BG9. In senescent second leaves, components **i**–**n** were significantly increased.

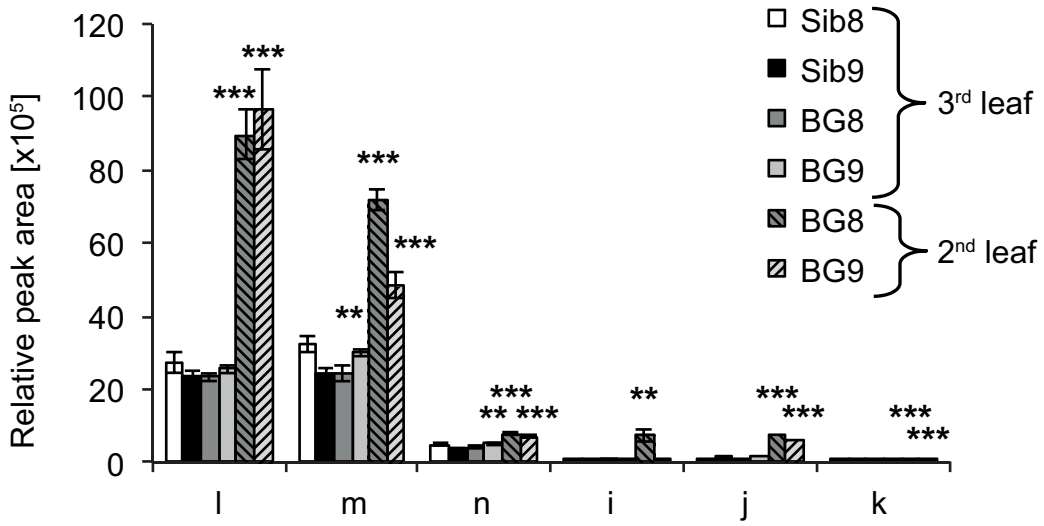


Figure 10: Relative quantification of components **i**-**n** in barley seedlings based on $[M+2H]^{2+}$ EICs. Peak areas are given relative to internal standard ampicillin, values are means and error bars represent standard error (* *p*-value ** < 0.01 and *** < 0.001 based on Student's *t*-test).

Relative isotope distributions revealed doubly charged ions of nitrogen containing molecules (Table 10), and according mass errors were generally below 5 ppm. Fragmentation obtained in (+)-ESI-MS/MS spectra revealed a guanidyl group, but not much information about the structure of the metabolites associated to components **i**–**n** could be obtained at this point.

The crucial hint for structure elucidation was received from transcriptomics experiments, which were performed by collaborators within the multi-disciplinary *Lr34* project. Microarray experiments were performed to compare the mRNA levels of transgenic *Lr34* barley plant with control plants [27]. Among the differentially expressed genes, many were related to biotic stress and secondary metabolism. The induction of key genes in phytoalexin biosynthesis was validated by be a more sensitive method, quantitative real-time polymerase chain reaction (qRT-PCR). Amongst the regulated transcripts, the gene of agmatine coumaroyltransferase was strongly upregulated in transgenic *Lr34* barley plants of BG8 and BG9 lines, especially in mature leaves (Figure 11)[27].

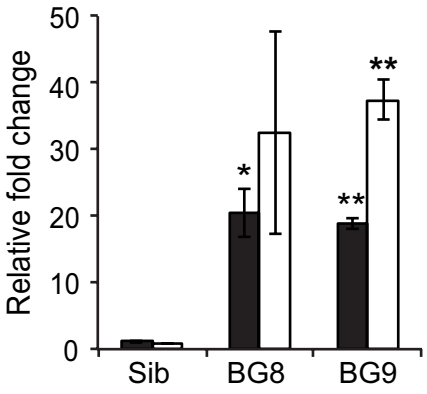


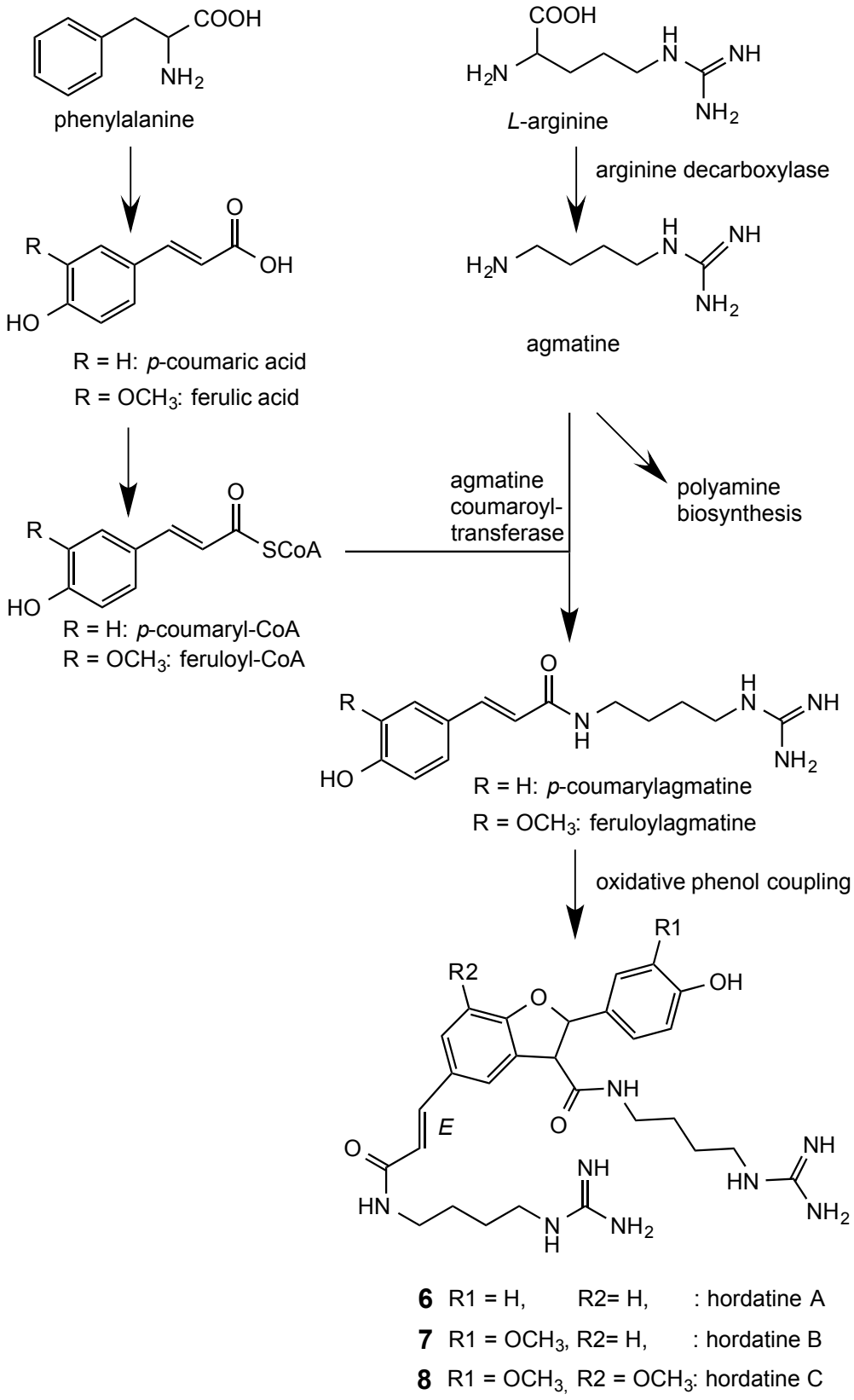
Figure 11: mRNA expression of agmatine coumaroyltransferase was strongly upregulated in BG8 and BG9 barley plants in both seedling (black) and mature leaves (white). Values are mean relative fold changes versus Sib and error bars represent standard error (* *p*-value <0.05, ** < 0.01 based on Student's *t*-test), based on [27].

The enzyme agmatine coumaroyltransferase has a high specificity for agmatine as acyl acceptor and preference for *p*-coumaroyl-CoA or feruloyl-CoA as acyl donor [117] and is involved in the biosynthesis of hordatines (Scheme 3), which are barley-specific secondary metabolites produced in seeds [118] and leaves. The name of these compounds is derived from a combination of *Hordeum* and agmatine.

Basically, oxidative dimerization of hydroxycinnamic acid amines conjugating two *p*-coumaroylagmatines result in hordatine A (**6**), conjugation of *p*-coumaroyl- and feruloylagmatine in hordatine B (**7**), and conjugation of two feruloylagmatines in hordatine C (**8**), respectively.

Hordatines function as phytoalexins, compounds that exhibit anti-fungal properties [119], and their concentration is known to slowly increase in barley leaves upon powdery mildew infection [120].

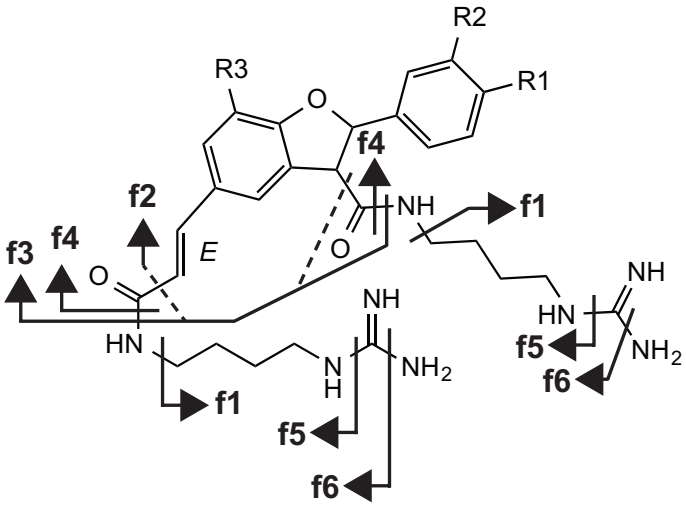
In the subsequent section, we examined hordatine derivatives as possible metabolites corresponding to components **i**–**n** and compared the measured MS/MS spectra with references from literature.



Scheme 3: Biosynthesis of hordatines according to [117] and [74].

Three of the increased components observed in Exp1 correlated with hordatines. Based on the calculated molecular formulas and on MS/MS fragmentations, component **I** was

correlated to hordatine A (**6**), component **m** to hordatine B (**7**) and component **n** to hordatine C (**8**). The (+)-ESI-MS/MS spectra of $[M+H]^+$ of components **l**–**n** displayed in Figure 12 characteristically showed sequential losses of an amino group, a guanidyl group, agmatine side chain and both agmatine side chains (Scheme 4). The neutral losses of 130 Da and 156 Da represented rupture of agmatine and agmatine with preserved amide bond, respectively. The mass shifts of Δ *m/z* + 30 Da between the fragments f4–f6 in the spectra of components **m** versus **l** and **n** versus **m** corresponded to methoxyl substituents. The MS/MS spectra of components **l**–**n** showed high similarity with spectra of the hordatines A, B and C described in barley grains [118] and in leaves [74]. Our metabolite annotation was based on comparison with literature data, thus corresponding to metabolite identification level 2 (chapter 1, 2.6).



6	R1 = OH,	R2 = H,	R3 = H	:	hordatine A
7	R1 = OH,	R2 = OCH_3,	R3 = H	:	hordatine B
8	R1 = OH,	R2 = OCH_3,	R3 = OCH_3	:	hordatine C
9	R1 = O-glucoside,	R2 = H,	R3 = H	:	hordatine A glucoside
10	R1 = O-glucoside,	R2 = OCH_3,	R3 = H	:	hordatine B glucoside
11	R1 = O-glucoside,	R2 = OCH_3,	R3 = OCH_3	:	hordatine C glucoside

Scheme 4: Proposed fragmentation of hordatines; f6 loss of amino group NH_3, f5 loss of guanidyl group CN_2H_2, f4 loss of side chain $C_5H_{14}N_4$, f3 loss of both side chains 2x $C_5H_{14}N_4$, f2 loss of both side chains $C_{11}H_{26}N_8O$, f1 side chain $[C_5H_{12}N_3]^+$.

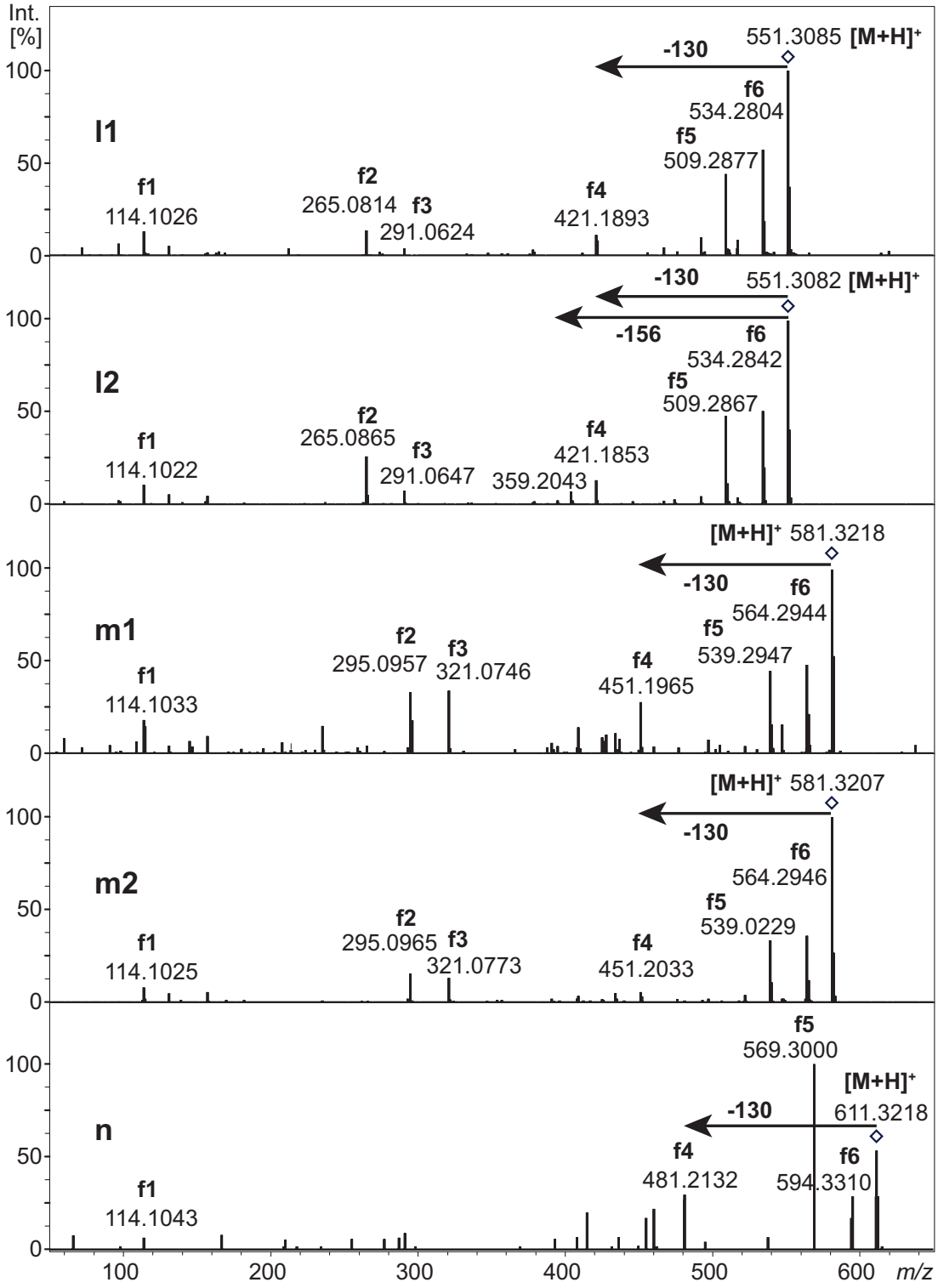


Figure 12: (+)-ESI CID-MS/MS of $[M+H]^+$ of components **l**–**n** at CE 35–70 eV, with fragment interpretations according to **Scheme 4**.

Hordatine glucosides were previously identified by NMR in barley seedlings [119] but also in grain and malt [118, 121, 122]. Accordingly, components **i**–**k** were annotated as potential hexose-derivatives of hordatines A, B and C (**9**, **10** and **11**, respectively) with mass shifts of Δ *m/z* 162 Da to the respective non-glycosylated variant. The only hydroxyl

group present in the hordatine molecules was the proposed site of glycosylation [119]. Due to the low abundance of hordatine glycosides, the isolation of precursor ions to generate MS/MS fragmentation spectra was not successful.

Table 11: List of hordatines with putative assignments, with retention times of chromatographic methods using a short 50 mm and a longer 150 mm column.

Com-ponent	Putative assignment	Molecular formula	MW [g mol^{-1}]	RT 50 mm [min]	RT 150 mm [min]	Reference in barley
i	Hordatine A glucoside **9**	$C_{34}H_{48}N_8O_9$	712.8	1.95 / 2.03	5.63 / 6.10	[74, 118, 123]
j	Hordatine B glucoside **10**	$C_{35}H_{50}N_8O_{10}$	742.8	1.90 / 2.02	5.35 / 6.07	[74, 118, 123]
k	Hordatine C glucoside **11**	$C_{36}H_{52}N_8O_{11}$	772.9	1.99 / 2.08	6.42 / 6.49	[74, 118, 123]
l	Hordatine A **7**	$C_{28}H_{38}N_8O_4$	550.7	2.23 / 2.35	7.22 / 7.89	[74, 118, 123]
m	Hordatine B **7**	$C_{29}H_{40}N_8O_5$	580.7	2.19 / 2.32	6.92 / 7.85	[74, 118, 123]
n	Hordatine C **8**	$C_{30}H_{42}N_8O_6$	610.7	2.21 / 2.37	7.34 / 8.29	[74, 118, 123]

For all hordatines corresponding to components **i**–**n**, double peaks were observed (Figure 8), with MS/MS spectra of high resemblance for components **l1**, **l2** and **m1**, **m2**, respectively (Figure 12).

Hordatines had been described as facile to *trans* / *cis* isomerization of the cinnamic acid moiety with varying amounts of *cis*-isomers [119]. Furthermore, stereoisomers of hordatine A with *cis*- and *trans*- configuration of two adjacent chiral carbons at the 5-membered ring had been reported in beer [124]. Based on the current data, the exact structure of the isomers could not be elucidated. However, a hint could be obtained from a previous study of barley seedling extracts applying similar chromatographic conditions as in the current thesis. Nomura and coworkers observed *cis*-cinnamic acid isomers of hordatines A and B to elute prior to the respective trans-isomers [125]. As a consequence, the peaks with earlier retention times (**l1**, **m1**) were assumed to be the *cis*- cinnamic acid isomers.

In a biological context, hordatines were long known to be phytoalexins with antibiotic activity against pathogens [119, 120, 126, 127]. In barley, increased hordatine concentrations were reported after powdery mildew infection [60]. The functions of hordatines were suggested to involve cell wall fortification to restrict pathogen invasion [117] and cytotoxicity towards the pathogen [128].

3.3 Hydrolysis of barley extract for the detection of apigenin and genkwanin

Following transcriptomics experiments performed by collaborators, the induction of selected genes was verified by qRT-PCR, including a strongly upregulated signal corresponded to the enzyme flavonoid 7-*O*-methyltransferase (F7OMT) (Figure 13). This *O*-methyl-transferase was previously reported to accumulate in barley leaves in response to attack of the pathogenic fungus *Blumeria graminis* causing powery mildew [129].

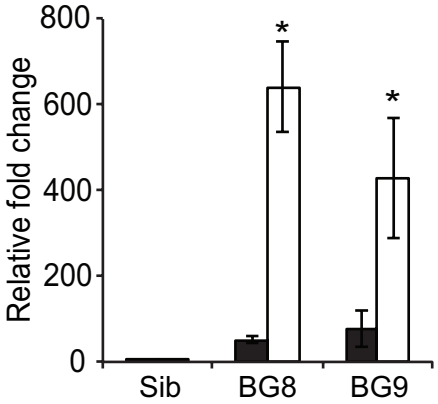
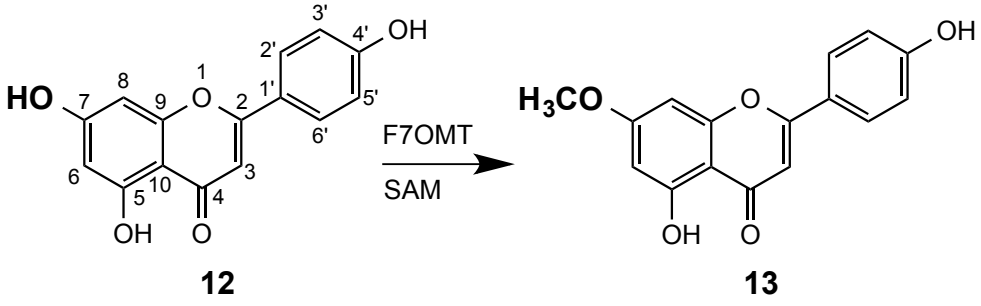


Figure 13: mRNA expression of F7OMT was strongly upregulated in BG8 and BG9 barley plants in both seedling (black) and mature leaves (white). Values represent relative fold changes versus Sib, error bars are standard error, significance * *p*-value < 0.05 was calculated by Student's *t*-test, based on [27].

The enzyme catalyzes the methylation of flavonoids at the 7-*O*-position by using *S*-adenosyl-*L*-methionin (SAM) as methyl donor, and its highest activity *in vitro* was shown for the substrate apigenin [129]. The methylation of apigenin (**12**) leads to genkwanin (**13**) (Scheme 5).



12 **13**

Scheme 5: F7OMT catalyzes the transfer of a methyl group from SAM to the hydroxyl group at the 7-*O*-position of apigenin (**12**), resulting in genkwanin (**13**).

Because F7OMT was strongly accumulated transgenic *Lr34* barley plants, we investigated apigenin and genkwanin, substrate and product of the according biotransformation, in metabolic extracts.

Therefore, reference compounds of apigenin and genkwanin were analyzed with the same chromatographic method used for the metabolic fingerprinting described in the previous sections (3.1, 3.2) to obtain the according retention times (apigenin $[M+H]^+$ 271.061 at RT 3.43 min, genkwanin $[M+H]^+$ 285.0712 at RT 4.29 min). Then, the data from Exp1–Exp4 were searched for the corresponding features. Neither apigenin nor genkwanin was observed in barley seedling extracts of the metabolic fingerprinting experiments Exp1–Exp4.

Generally, free flavonoid aglycones are far less abundant in plants than their water-soluble glycosides [130]. The glycosylation variety for apigenin aglycone is considerable as a result of booth *C*- and *O*-glycosylation, different substitution positions, and different types of sugar moieties attached. Apigenin derivatives frequently include 8-*C*, 6-*C* [131] as well as 7-*O*-glycosides [132]. Analysis of different glycosylations can be achieved by MS/MS for screening and identification of acylated flavonoid-*O*-glycosides and methoxylated flavonoids [133].

Another approach to simplify the analytical procedure is to release aglycones in plant extracts by hydrolysis, and had been described for glycosylated flavonoids e.g. in spinach [134] or *Equisetum* [135].

For the hydrolysis of barley plant material, a method with acidic conditions was chosen that was tested with the flavonoid-*O*-diglycoside hesperidin. Four samples were prepared for the hydrolysis experiment: a sample of reference compounds, a sample of plant material of the sister line spiked with reference compounds, a sample of plant material of the sister line and a sample of plant material of the of transgenic BG9 line. The reference mixture contained each 50 ng ml^{-1} apigenin (**12**) and genkwanin (**13**).

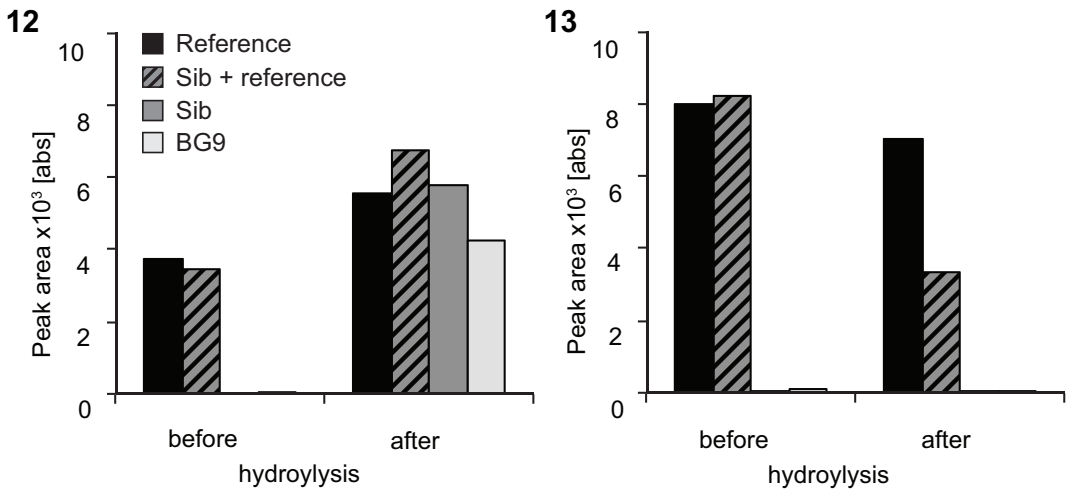


Figure 14: Hydrolysis of plant material standard addition: reference compounds apigenin (**12**) and genkwanin (**13**).

Figure 14 shows the signal intensities of apigenin (**12**) and genkwanin (**13**) detected before and after hydrolysis in the four samples. For apigenin, the signal in the reference sample (black) was higher after hydrolysis than before due to experimental variation. After hydrolysis, apigenin was detected in hydrolysates of sister line (dark grey) and BG9 (light grey) at around the same concentration as in the reference sample. Due to the experimental variation observed for the reference sample, no clear difference between sister line and BG9 line was observed. For genkwanin, no signal was detected in hydrolysates of sister line and BG9.

So, despite the accumulation of the enzyme F7OMT, none of the methoxylated product of alanine, genkwanin, could be observed. In further hydrolysis experiments performed for alternative substrates of F7OMT that were reported to be methoxylated with a lower relative enzymatic activity [129], naringenin and apigenin-6-*C*-glucoside (vitexin), also no methoxylated products could be detected after hydrolysis (Supplementary scheme 1, Supplementary figure 3).

Thus, no methoxylation effect of the most prominent substrates of F7OMT, was observed in *Lr34* plants.

3.4 Discussion

To sum up, metabolic fingerprinting experiments to characterize transgenic *Lr34* barley were performed by LC-MS. Among the differentially regulated metabolites, a group of increased cyano glycoside derivatives in experiment Exp3 and another group of upregulated hordatine derivatives in experiment Exp1 was identified based on HR-MS and MS/MS data. The increase in concentrations of hordatine derivatives in *Lr34* plants compared to control plants biochemically correlated with the increased mRNA levels of agmatine coumaroyl transferase, a key enzyme in the hordatine biosynthesis pathway detected in transcriptomics experiments [27]. In the other experiments, these two groups showed a tendency towards increase, but only with partial statistical significance. Between the experiments, different conditions were changed, including plant growth conditions, plant age at harvest and extraction procedure (Table 4). This might possibly cause the different experimental outcomes.

Another gene that was observed to be upregulated in *Lr34* barley, encoding a flavonoid-7-*O*-methyltransferase, was found to have no effect on the methoxylation of its most prominent substrate, apigenin.

Variations in cyano glycoside and hordatine derivatives concentrations

Cyanogenic glycoside concentrations were reported to show considerable variation, even in individual plants of genetically identical cassava and white clover [115]. Other factors proclaimed to influence cyanogenic glycoside concentrations were plant development (ontogeny), climate (phenology) and time (chronology) [115]. Generally, higher cyano glycoside concentrations were observed under growth-limiting environmental conditions, and abiotic factors reported to elevate cyano glycosides involve salt stress (in white clover [136]), drought (in sorghum and eucalyptus [115]), elevated temperature and ozone (in barley seedlings [60]), possibly to alleviate oxidative stress [137].

Transgenic *Lr34* barley plants develop severe premature leaf-tip necrosis, and plant growth and development lack behind the control plants. In the context of the observations described above, growth delayed sorghum plants were described as phenologically younger, displaying higher cyano glycoside concentrations [115, 138]. This corresponds with the general trend that cyano glycoside concentrations are highest in seedlings and decrease with age [115]. Accordingly, cyano glycoside accumulation in transgenic *Lr34* barley might be an indirect effect caused by the growth retardation.

In terms of sample preparation for hordatine analysis, the extraction with a slightly more concentrated methanolic extraction solvent in Exp1 versus Exp2–Exp4 was not expected to have a big influence, as extraction procedures with varying methanol and acid content were described [118, 123, 127] [139]. Therefore, the variation was expected to arise from biological factors.

One factor influencing hordatine levels in barley was the developmental stage of the plants, and previous studies reported a decrease of hordatine with barley seedling age. The maximum hordatine concentration in barley seedlings was reported 2–6 days after germination [139, 140], with a decline to 50% after 10 days, before reaching stable 10% of the maximum hordatine concentration after 28 days. In contrast, we observed higher absolute hordatine concentrations in older plants of Exp1 compared to younger plants of Exp2–Exp4. For Exp1, the developmental stage of the barley seedlings at age of 21 days was more advanced than after 14 or 16 days (Exp2–Exp4).

Additionally, also abiotic factors affect hordatine accumulation. Hydroxycinnamic amides, which are precursors of hordatines, were reported to change in barley under abiotic stress, e.g. upon osmotic stress [141] or at elevated temperature [60]. For barley seedlings grown under elevated temperature (24/17 °C versus 19/12 °C day/night), strongly enhanced levels of hordatines were observed [60]. Also combined treatments of both elevated temperature and CO_2 (700 ppm versus 485 ppm) or additional exposure to O_3 (100 ppb) lead to hordatine accumulation [60]. Another study correlated ecogeographic and climatic characteristics of the parent plants indirectly with hordatine levels in seedlings of 50 wild barley accessions, and temperature was found to negatively correlate with hordatine concentrations [139].

In the current study, barley plants of experiment Exp1 — which showed most significant accumulation of hordatines in *Lr34* plants — were grown in the greenhouse, whereas Exp2–Exp4 were grown under controlled conditions in a phytotron chamber. Considering the strong effects of abiotic factors on hordatine accumulation reported in literature, we assumed that the greenhouse conditions used in Exp1 supported the increase of hordatine derivatives in *Lr34* plants.

Possible functions of cyano glycoside and hordatine derivatives in barley

The function of cyano glycoside derivatives in barley in general and also in connection with *Lr34* is not fully clear. Cyanogenic glucosides were reported to be multifunctional secondary metabolites with biological roles in defense, nitrogen storage and transport,

seed dormancy and germination, and potentially in regulation of abiotic stress and ROS [116].

A general theory about resource allocation assumes that the synthesis and maintenance of defensive compounds demands a cost for the plant, provoking sacrifice in terms of plant growth or reproductive output [115]. For cyanogenic glycosides, these costs were difficult to measure and reported to be low in terms of the overall energy budget [115].
Because cyanogenic glycosides evolutionarily acquired roles in other metabolic pathways, the energy cost might be compensated. The synthesis of cyano glycosides was even suggested to be a strategy to dissipate excess energy and reducing power in terms of NADPH, which in turn was mitigating stress [115].
Metabolic turnover and remobilization of resources is another strategy how costs of plant secondary metabolites can be balanced. The nitrogen and sugar content of cyano glycosides may represent significant percentages of the total organic matter of a plant organ. Upon requirement, the plant can remobilize or reallocate the nitrogen and carbon deposited in cyano glycosides to cover resource demands e.g. in primary metabolism. An example for metabolic turnover of cyanogenic glycosides is found in the rubber tree, where linamarin accumulates in seeds and is glycosylated into diglucoside linustatin upon germination. The diglucoside is then cleaved into cyanohydrin and further transformed into asparagine. Alternatively, sequential glucoside cleavage releases the initial linamarin, which can be either stored in cotyledons as defense chemical or relocated into young leaves and roots. Linustatin is also mobilized to the bark of rubber tree, where carbon units are acquired for rubber production [116]. This example illustrates the multiple functions of cyanogenic glycosides and the flexibility of metabolic turnover in plants. Also in barley, especially in the epidermal cell layer, a high percentage of the total soluble carbohydrate content is bound to cyano glycosides [114]. Therefore, the functions of cyano glycosides in barley, which are not yet fully understood, might well involve storage of nitrogen and sugar for metabolic turnover. In transgenic *Lr34* plants, the accumulation of cyano glycosides might be a temporary intermediate for remobilizing resources for other plant defense responses.

Hordatines function as phytoalexins and were associated with cell wall fortification upon pathogen invasion. Accumulation of hydroxycinnamic acid amides derived from the phenylpropanoid pathway had been described as a stress response in barley, wheat, oat, potato, tobacco, tomato and carnation [117]. The most prominent increase in hordatines

was observed in senescent secondary leaves of Exp5, thus suggesting being part of a general stress response.

Metabolic context of *Lr34*

Besides the increase of cyano glycoside and hordatine derivatives, we observed several constitutive defense responses in barley as part of the multi-disciplinary project to unravel the *Lr34* resistance mechanism [27]. Both lignin content and expression of the respective biosynthetic genes increased in transgenic *Lr34* lines. Lignification at the site of infection is a known important defense mechanism against penetrating fungal pathogens. In this respect, elevated lignin levels likely contribute to the constitutive, basal plant resistance by *Lr34*. Furthermore, raised levels of tryptophan and phenylalanine were observed in *Lr34* plants, which are precursors for a multitude of secondary metabolite pathways such as phenylpropanoids, flavonoids, alkaloids, lignins and phenolic polymers. Much higher amounts of the plant hormones salicylic acid and jasmonic acids were found in transgenic *Lr34* barley leaves, suggesting a broad general defense response.

The maintenance of multiple defense responses over a considerable time period requires a high energy demand, and increased defense-related pathways were compensated by reduced photo- and chlorophyll synthesis [142]. In fact, transgenic *Lr34* barley plants show severe constraints in growth and development in combination with premature leaf-tip necrosis while maintaining resistance against multiple pathogens. In contrast, this fitness penalty was not observed in wheat, where *Lr34* showed only minor effects on grain yield [30]. While wheat contains an additional susceptible *Lr34* allele, the barley genome includes no ortholog of *Lr34* [22]. Thus, the susceptible allele might influence the activity of *Lr34* in a yet unknown manner.

4 Metabolomic profiling of *Lr34* resistance in barley, rice and wheat by LC- and GC-MS

The non-targeted metabolic fingerprinting approach described in the previous chapter is a versatile tool to detect a high number of metabolites, but the applied protocol has intrinsic methodological limitations that restrict the detectable metabolic range to a limited part of the whole metabolome, with focus on secondary metabolites. To be able to observe a broader range of plant metabolites, a combination of different extraction solvents and orthogonal detection techniques had been applied to investigate *Lr34*-mediated resistance by a metabolic profiling approach described in this chapter.

The purpose of this extensive investigation, performed in collaboration with the Max Planck Institute of Plant Physiology (MPIMP), was to broaden the metabolic coverage by combining complementary methods based on LC- and GC-MS and to detect primary and secondary metabolites as well as lipids and selected plant hormones in barley, rice and wheat, grown under different conditions. This approach allowed gaining a more complete picture of the metabolic pathways involved in *Lr34*-mediated resistance.

4.1 Design of the investigation

The major aim of the experiment was to gain a deeper insight into the metabolic pathways related to *Lr34* resistance by characterization of differences between *Lr34* plants versus control plants. Further aims were to investigate differences in the plants' defense responses upon infection, and to compare metabolic profiles across the cereal species barley and rice. In order to account for experimental and biological variation, the experiments for each condition were performed twice. For barley and rice, each two genotypes with low and high expression levels of *Lr34* were investigated to bypass effects of the leaf-tip necrosis phenotype. In a further step, field-grown wheat was investigated (4.7).

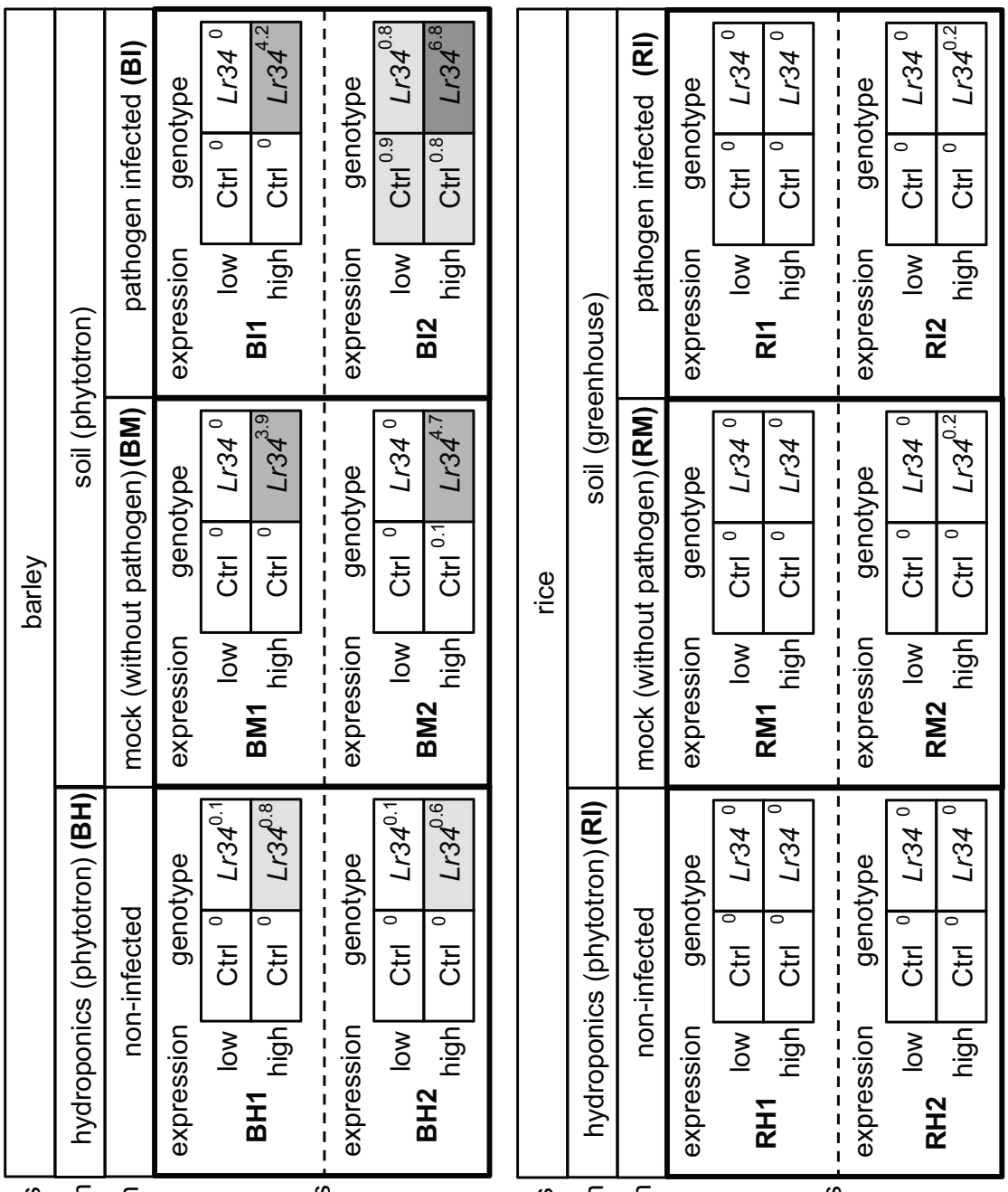


Figure 15: Design of the investigation for barley (A) and rice (B). In each experiment, the metabolomes of *Lr34* plants were compared with respective control (Ctrl) plants. Gray scale and numbers represent the estimated senescence in the harvested leaf (senescent part of the leaf relative to leaf length scaled to 10).

All experiments were designed to compare the metabolomes of *Lr34* plants with control plants, in order to find significant differences caused by the resistance gene (Figure 15). As controls for each transgenic *Lr34* line of barley and rice, sister lines in which the construct segregated out of the respective line were utilized. Two genotypes with low and high expression levels of *Lr34* for both barley and rice were used, resulting in low- and high number of transcripts and respective transporter proteins. Even though *Lr34* is expressed with lower abundance in low expression lines, the resistance is still functional. Lower expression levels are generally correlated with less phenotypic effects and show less leaf-tip necrosis, and the senescent part of the leaves was estimated by scaling the dry part of the harvested leaf relative to total length of 10 (Supplementary figure 4). All experiments in barley and rice were performed in twice, resulting in two batches per growth condition.

For the species barley and rice, three well defined, different experimental conditions were used (Figure 15). For the first growth condition, barley and rice were grown on hydroponic solution as growth medium and in a phytotron chamber (abbreviated as BH and RH for barely and rice). Plants grown under standardized conditions — including a well defined hydroponic growth medium in a phytotron with controlled temperature and light regime — were hypothesized to lead to less biological variation within a group of replicates, and thus to a higher number of significantly changed features between *Lr34* and controls. Accordingly, the metabolomes of hydroponically grown plants were expected to be more robust than the other growth conditions, which would allow a better characterization of the metabolites involved in *Lr34*-mediated resistance. To investigate differences in the plants' defense responses upon infection, plants without pathogen (mock infected, BM and RM) and plants infected with fungal pathogens (BI and RI) were grown on soil. The mock infection condition was implemented by spraying blank solution without fungal spores to plants of the same growth stage, while no spray was applied for the hydroponic growth condition. For the infected growth condition, barley plants were infected with the biotrophic barley leaf rust (*Puccinia graminis sp hordei*) and harvested after clear infection symptoms were recognized (8 and 11 days post infection for BI1 and BI2 respectively). Rice plants were infected with the fungal pathogen rice blast (*Magnaporthe griseae*), and leaves were harvested after 28 hours post infection. This early time point was chosen because *Lr34* was shown to provide resistance against biotrophic fungi, and rice blast is a hemi-biotrophic fungus that proliferates in a biotrophic manner the first 72 hours of infection [9], before the pathogen adopts a necrotic lifestyle later on [143] To compensate for variations in plant growth, sample preparation and analytical methods, sets of six replicate plants

were selected for each genotype of barley and rice. Details about plant material are documented in experimental procedures 6.2.1.

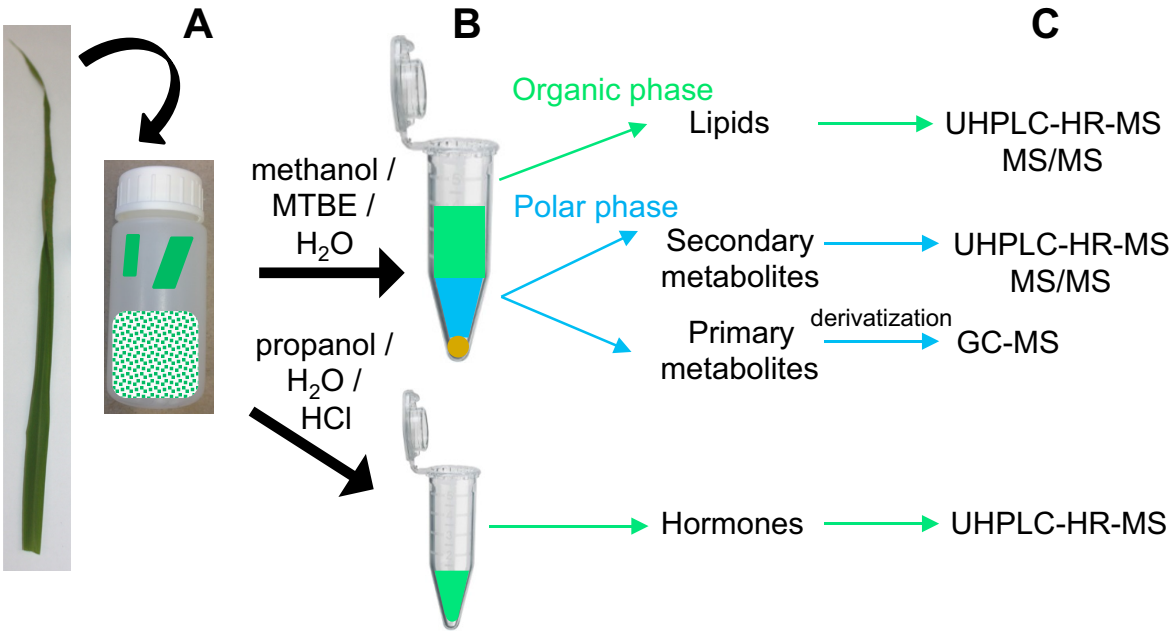


Figure 16: Sample preparation of primary, secondary metabolites lipids and hormones; homogenization of whole leaf (A), fractionated extraction of plant powder (B), analysis method for individual fractions (C).

For all samples, whole leaves were harvested and homogenized (Figure 16A). In order to enlarge the metabolic coverage, the plant powder was extracted by a mixture of methanol, methyl-tert-butylether (MTBE) and water (Figure 16B). This extraction protocol allowed the simultaneous extraction of primary and secondary metabolites as well as lipids from the same samples [144]. The apolar organic fraction was analyzed by UHPLC-HR-MS with a chromatographic method specific for lipids. The polar phase was split into separate aliquots for parallel analysis of secondary and primary metabolites by UHPLC-HR-MS and GC-MS, respectively. A two-stage derivatization procedure including methoxyamination and silylation was applied prior to GC-MS analysis in order to detect non-volatile, thermally instable primary metabolites such as monosaccharides, amino-, organic and fatty acids. Plant hormones were extracted separately and detected by UHPLC-HR-MS.
In summary, the experiments of the three species barley, rice and wheat consisted of 378 samples, resulting in over thousand analyses.

While primary metabolites and lipids were investigated by targeted metabolic profiling, non-targeted metabolic fingerprinting was performed for secondary metabolites.

After mass-spectrometry based data analysis, data were preprocessed to result in two-dimensional matrices comprising metabolite identities or feature information and intensities of all replicate samples. To account for variances that arose from sample preparation and metabolite analysis, data were normalized to the intensity of the internal standards and to fresh weight of the plant material. Furthermore, the matrices were filtered with the metabolic feature intensities in blank samples and further for occurrence, robustness and abundance. For the visualizations including principle component analysis (PCA) and heatmaps, the data were log2 transformed (experimental procedures 6.2.4).

Next, a univariate statistical analysis was performed, more specifically an analysis of variance (ANOVA) including multiple testing corrections and further evaluation by a post-hoc Tukey's test. The significantly different metabolites with *p*-values below 0.05 — reflecting a widely accepted significance threshold of 5% — were considered for comparison between independent experiments described below.

4.2 Metabolic profiling of primary metabolites

The metabolic profiling of primary metabolites was based on GC-EI-MS data that were processed with the *TagFinder* software [145]. For metabolite identification, the experimental spectra were compared with reliable spectra collected in the publicly available *Golm Metabolome Database* containing authenticated reference compounds including groups of amino acid derivatives, sugars and polyols, organic acids and fatty acids [146]. After manual verification of the spectra, 84 primary metabolites were identified in barley and rice (Supplementary table 3).
For each of the identified metabolites, a single fragment was selected to be used for quantification, and the respective intensities were summarized in a two dimensional matrix for statistical analysis in order to classify differences between the primary metabolic profiles of the control and *Lr34* plants.

To explore the characteristics of the dataset of the identified primary metabolites, a multivariate analysis was performed by principal component analysis (PCA, chapter 1, 2.5). Visualizations of two representative PCA score plots are shown in Figure 17 for mock-infected barley and rice. The strongest group separation was obtained for high-expressing barley, which was clearly separated from the high-expressing control group and the low-expressing groups. In rice, the group separation was less pronounced for high *Lr34* expression levels, and no clear group distinctions could be obtained for the low expression lines. While low expression samples were located around the center of the plot, highest variance could be observed for control and *Lr34* expression lines. These results indicated that the strongest discriminating power of the primary metabolite data was observed for high *Lr34* expressing barley, which also displayed a strongly pronounced senescence phenotype.

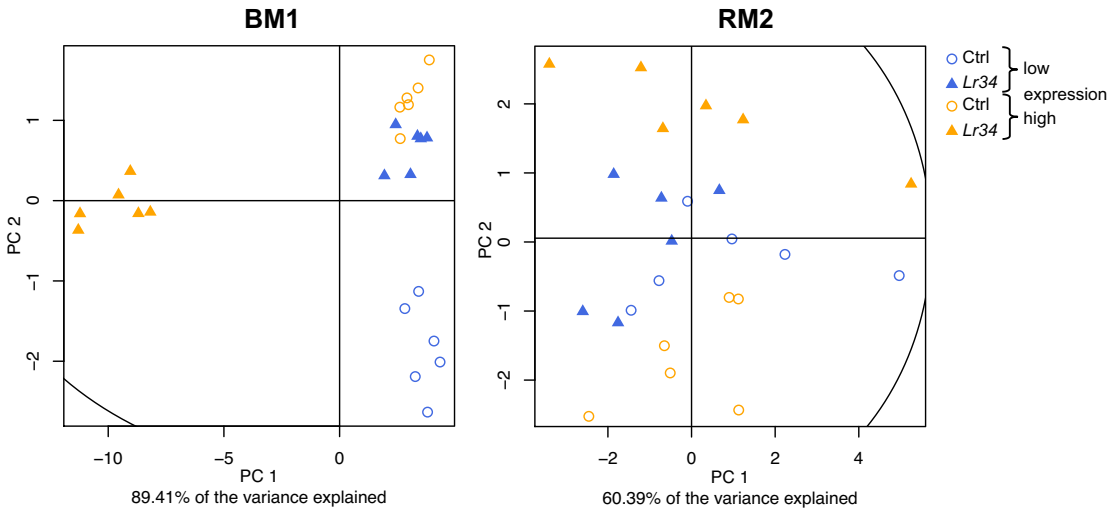


Figure 17: Representative PCA score plots of the identified primary metabolites in barley (left) and rice (right) grown under mock-infected condition. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles. A clear group separation was obtained in barley for both expression levels and for rice at high expression level, but not for low-expressing rice.

Then, the statistical relevance of each metabolite was evaluated by an analysis of variance (ANOVA) with multiple testing corrections and post-hoc Tukey's test, and metabolites with significant differences corresponding to *p*-values below 0.05 were compared in individual experiments.

Two different strategies were applied to highlight common metabolic differences observed in individual experiments, and they were based on a single species. The two different strategies are depicted in Figure 18 and exemplified for hydroponically grown barley. The first strategy (A) considers both low and high expression levels simultaneously for both batches of each growth condition. For example, metabolic differences of the four groups BH1 low expression, BH1 high expression, BH2 low expression and BH2 high expression were compared. Metabolites that were significantly regulated in all four groups (intersection marked with black cycle) and metabolites that were intersecting amongst three groups (marked with grey circle) were selected for the comparison between the three growth conditions hydroponics, mock and pathogen infected. In the second strategy (B) the two expression levels, low and high, were evaluated separately. For example, the significant metabolic differences in the two batches grown under hydroponic conditions, BH1 and BH2, were compared for plants with low expression levels of *Lr34*. The intersection of these hydroponic batches was than compared with the respective intersections of mock-infected and pathogen-infected experiments.

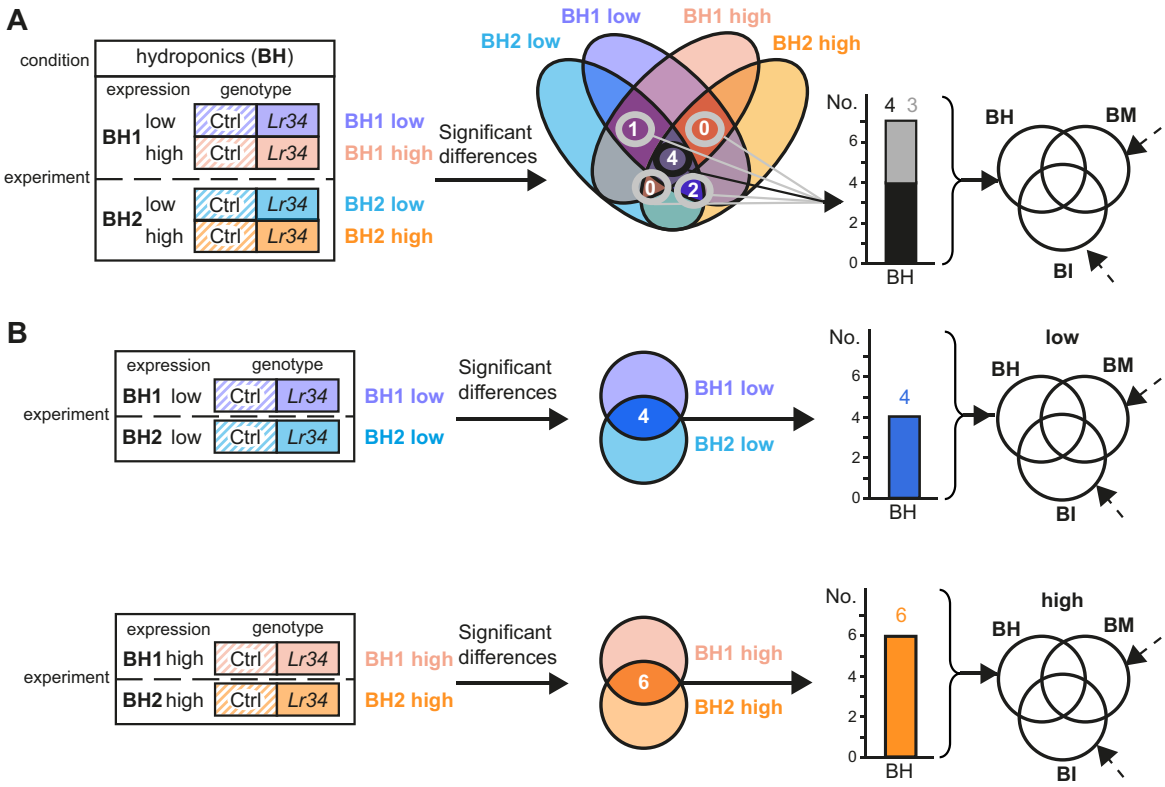


Figure 18: Common metabolic differences amongst independent experiments in barley and rice: (A) consideration of both expression levels (low, high) and both batches (e.g. BH1 and BH2); (B) consideration of low- and high expression level separately; intersections of significantly different metabolic signals was used for comparison between conditions. For (A), overlaps of 4 groups (black circle) as well as 3 groups (grey circles) were considered.

For the identified primary metabolites, the first strategy (A) that simultaneously considered both expression levels and both batches resulted in a small number of common differences in both barley and rice (Figure 19A). For barley, a single metabolite was robustly upregulated in three groups under hydroponic growth conditions, namely 2-aminoadipate. For the barley plants grown under mock-infected conditions, the concentration of 1-*O*-methyl-glucopyranoside was increased in all four groups (black) whereas glucose- and fructose-6-phosphate were upregulated in three groups (grey). For rice, isocitrate was consistently downregulated for plants grown under hydroponic conditions whereas citrate was upregulated in transgenic *Lr34* plants in grown under both mock- and pathogen-infected conditions. The subsequent comparison of the three different growth conditions hydroponics, mock- and pathogen-infected produced no intersection.

The second strategy (B) that separately evaluated the commonly regulated metabolites in low expression level, colored in blue, and high expression level, colored in orange, showed a similar result (Figure 19B). The increased concentration of 2-aminoadipate that was observed with the first strategy in hydroponically grown barley, and the upregulation of 1-*O*-methyl-glucopyranoside in mock-infected barley was also observed at low expression level. Furthermore, pyruvate and glycine were decreased in low expressing mock-infected barley. In low-expressing rice, isocitrate was consistently downregulated in plants grown under hydroponic condition whereas citrate was upregulated in transgenic *Lr34* plants with mock- and pathogen infection.

For the high expression level, more primary metabolites were commonly regulated. Hydroponically grown rice with high expression levels of *Lr34* showed increased oncentrations of succinate and galactinol, whereas concentrations of isocitrate, myo-inositol, allantoate, ascorbate and tyramine were decreased.

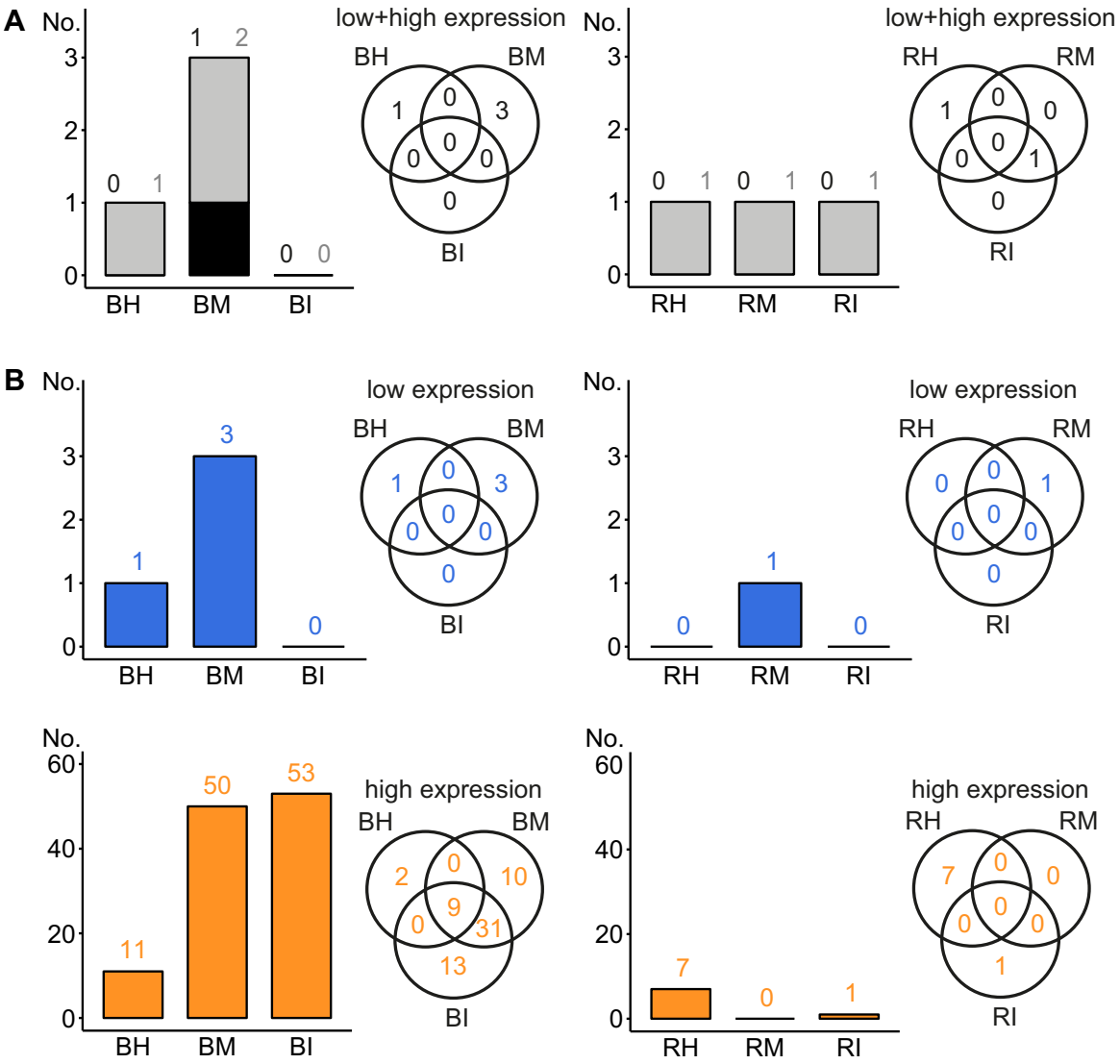


Figure 19: Commonly altered primary metabolites with same regulation tendency and FC > 0.3, based on: (A) combined low and high expression levels of barley and rice with overlaps of 4 and 3 groups (black and grey); (B) separate expression levels (low - blue, high - orange).

The highest intersection of primary metabolites (9) was detected in barley with high expression levels of *Lr34*. The intensities of nine metabolites were significantly altered under all conditions BH, BM and BI. In Figure 20, the log2 fold changes of each identified metabolite are displayed for all experiments in barley and rice. Red color indicates an increased metabolite level in *Lr34* versus control, whereas blue reflects a decreased metabolite level. Significant changes are marked with asterisks, and common metabolic changes in different growth conditions are highlighted with yellow frames. The nine metabolites that were regulated under all barley growth conditions are indicated with bold yellow frames and included the enhanced metabolites leucine, tyramine, organic acids succinate and γ-butyric acid (GABA), sugars fructose and glucose as well as putrescine.

Asparagine was not consistently regulated; while the asparagine concentration was decreased in plants grown on hydroponics, it was strongly increased under mock- and infected conditions. The metabolites regulated in high *Lr34*-expressing barley grown under all three growth conditions were mapped against a KEGG pathway map of rice (Supplementary figure 13). Furthermore, 31 primary metabolites were commonly upregulated in mock- and pathogen-infected high-expression lines of barley (Supplementary table 4).

The only consistently down-regulated metabolite in all BH, BM and BI was dehydro-ascorbate. In terms of biological interpretation, ascorbate could be important for the regulation of plant defense.

Upon pathogen infection, plants react with an oxidative burst at the infection site and release reactive oxygen species (ROS). ROS induce further antimicrobial responses and prevent the penetration through the cell membrane and spreading of biotrophic fungi through the plant, but high ROS concentration levels can damage cellular components. Scavenging systems orchestrate oxidative stress response mechanisms and thus regulate the complex ROS function in plant-pathogen interactions [147]. Amongst them, the ascorbate-glutathione cycle was demonstrated to play a pivotal role in the scavenging of reactive oxygen species. Dehydroascorbate and the radical monodehydroascorbate are recycled to ascorbate by the corresponding mono- and dehydroascorbate reductase enzymes. A monodehydro-ascorbate reductase was found to be regulated by miRNA in wheat resistance to stripe rust [148], and the concentrations of dehydroascorbate reductases have been shown to be induced upon infection of resistant rice with bacterial blight [149] and in resistant oilseed brassica crops with the necrotrophic fungi *Leptosphaeria maculans* [150]. In a previous study of *Arabidopsis thaliana,* infected with the necrotrophic ascomycete fungus *Alternaria brassicicola*, dehydroascorbate levels of infected leaves were significantly downregulated [151].

Putrescine and the polyamine degradation product γ-butyric acid might be involved in plant stress tolerance [99].

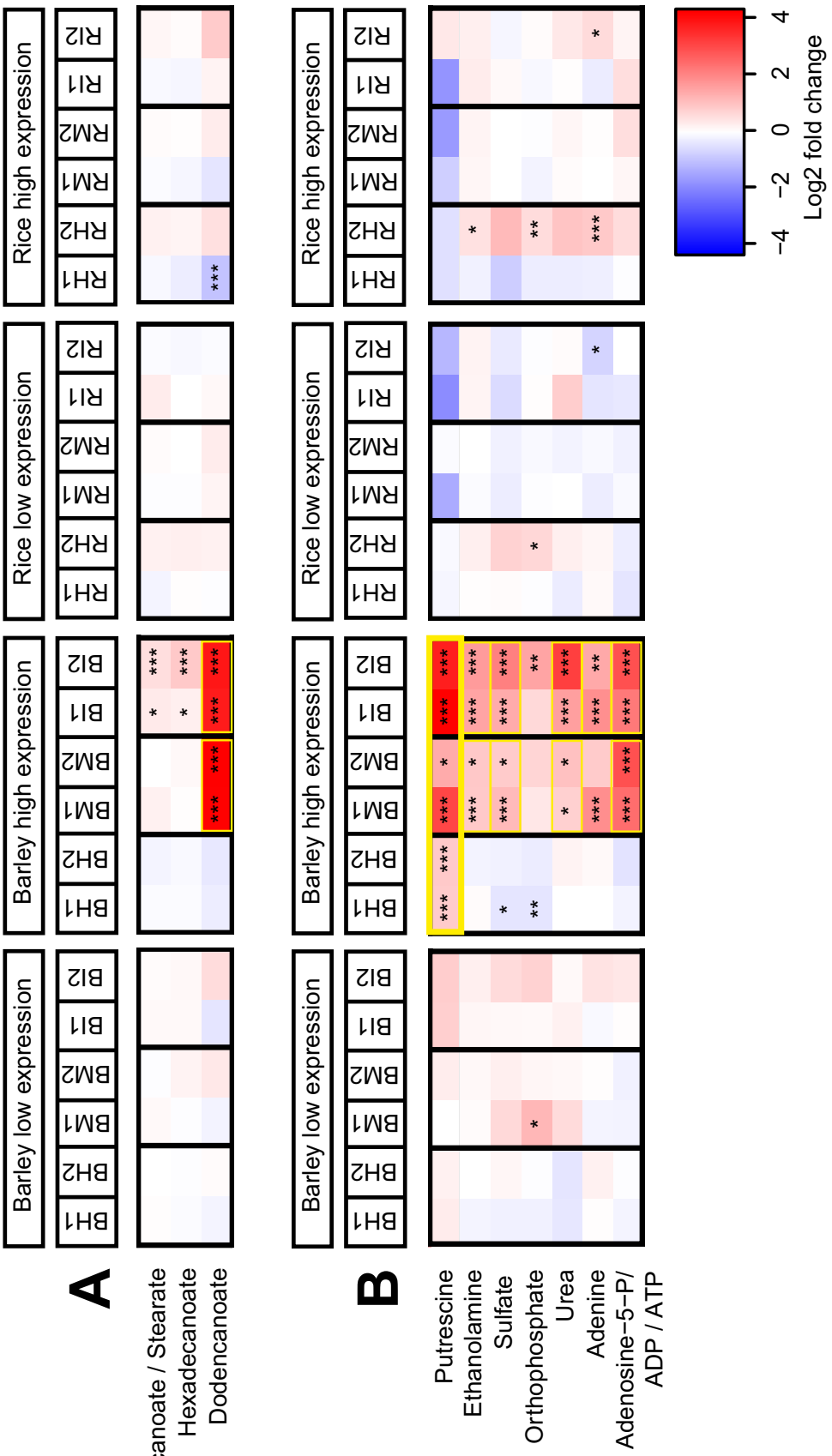


Figure 20: Heatmap of primary metabolic changes in barley and rice in different experiments, with commonly changed metabolites highlighted with yellow frame. Primary metabolites grouped in fatty acids (A), other primary metabolites (B), amino acids and derivatives (C), organic acids (D) and sugars (E). Ratios of log2-transformed fold changes are given by shade of red or blue according to the scale bar. Data represent mean values of six replicates. Statistical analysis was performed with ANOVA and post-hoc Tukey's test, of which * *p*-value <0.05, ** < 0.01 and *** < 0.001; grey asterisks indicate either fold change below 0.3 or inconsistency of regulation tendency.

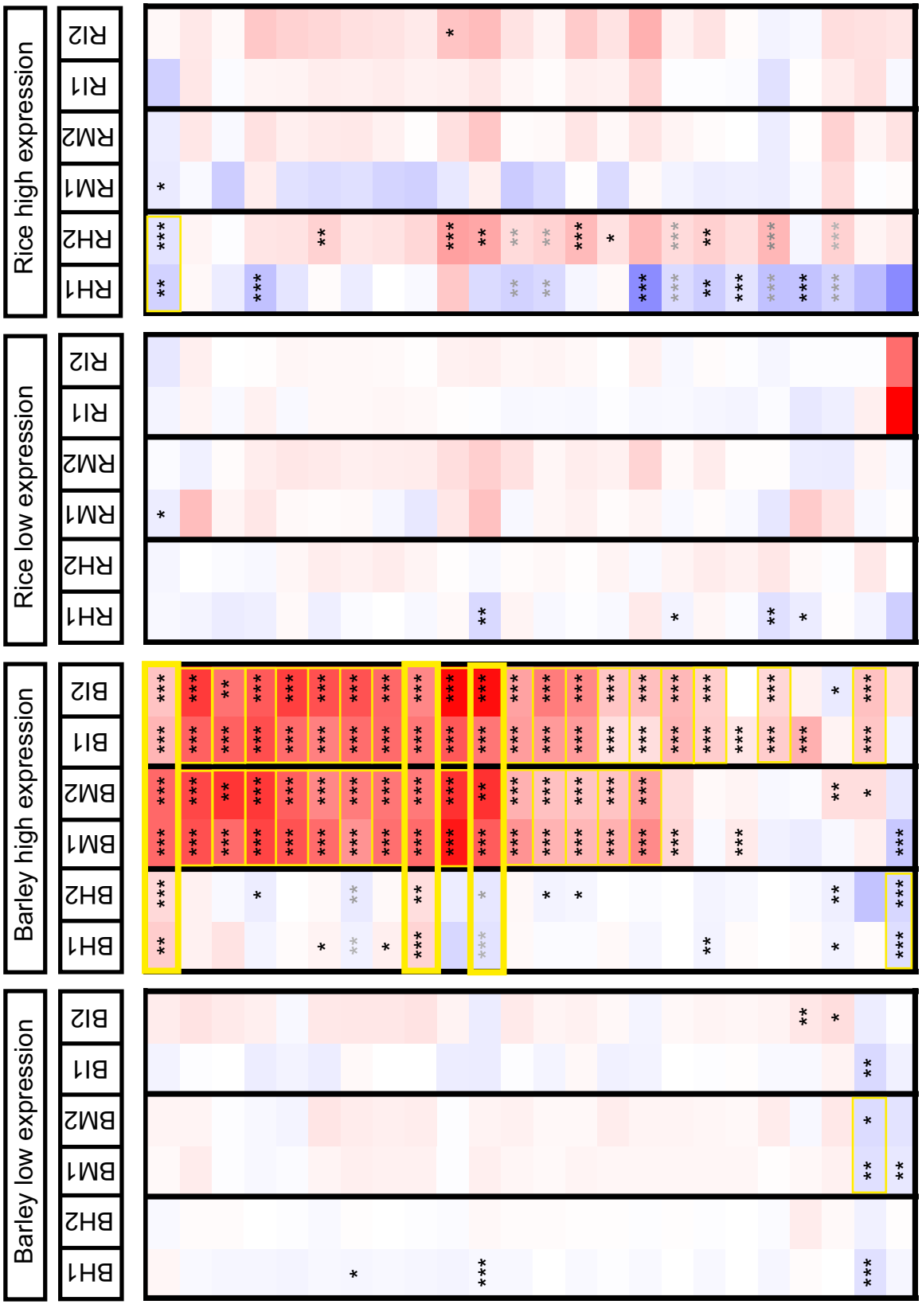


Figure 20, *continued*

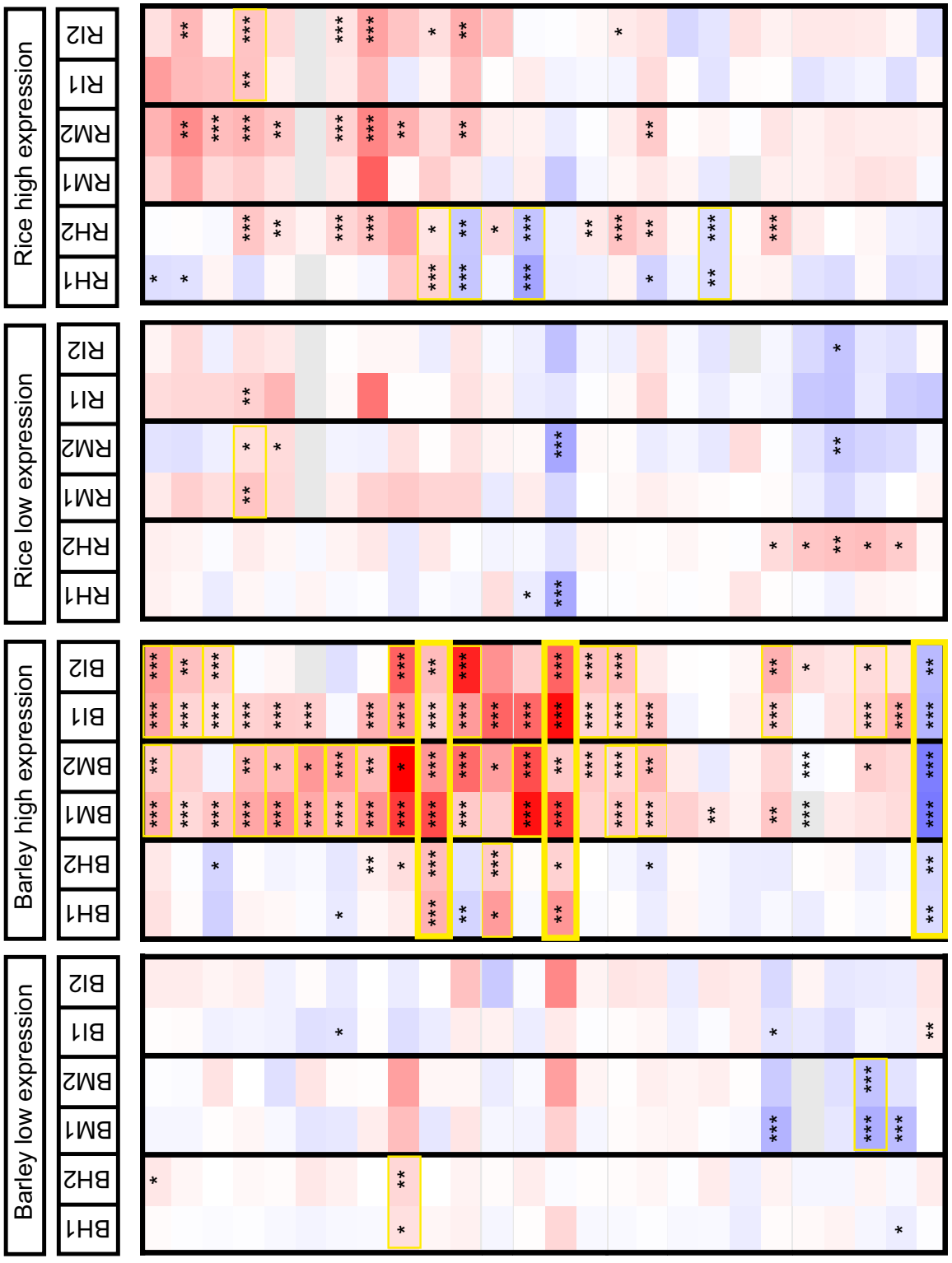


Figure 20, *continued*

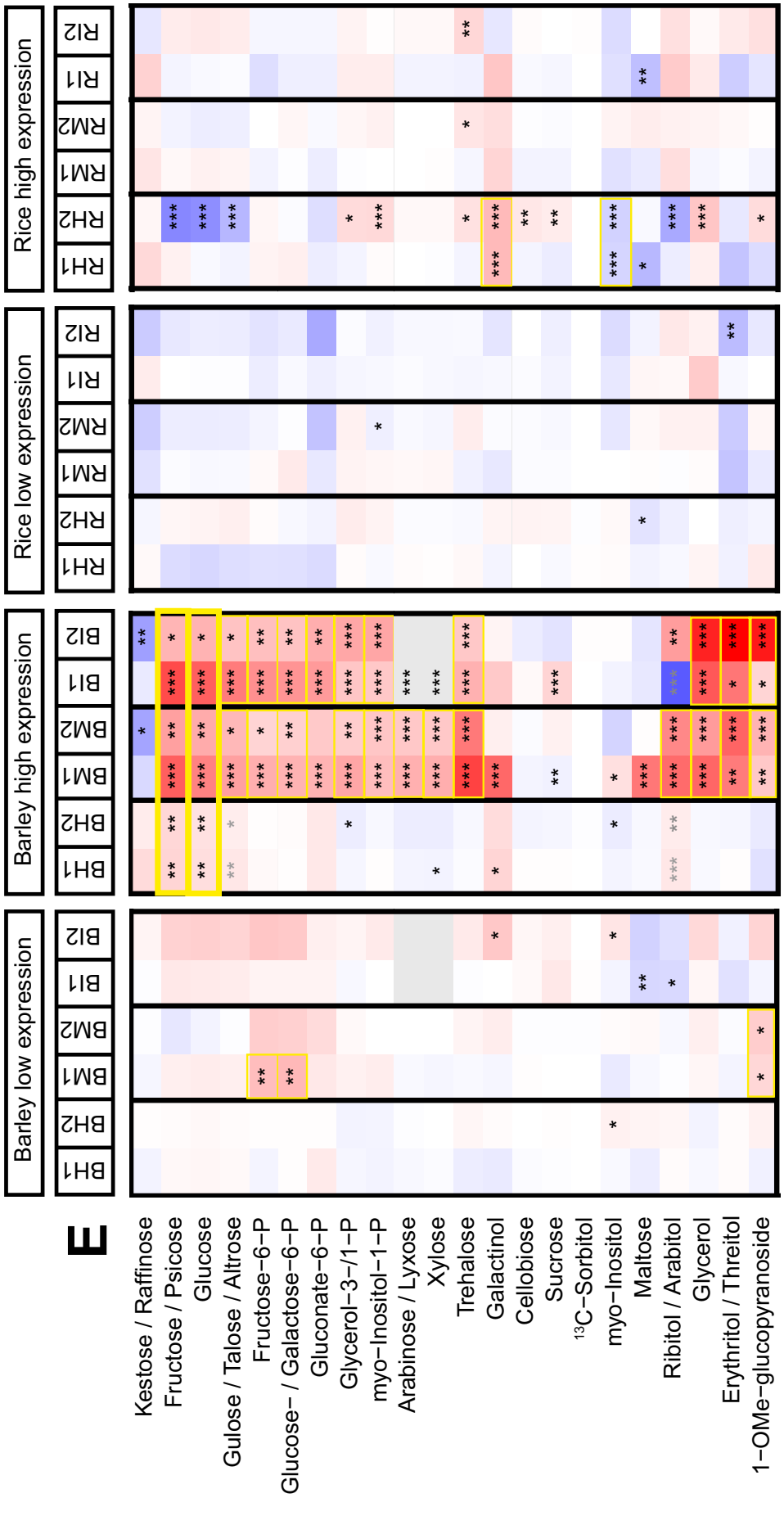


Figure 20, *continued*

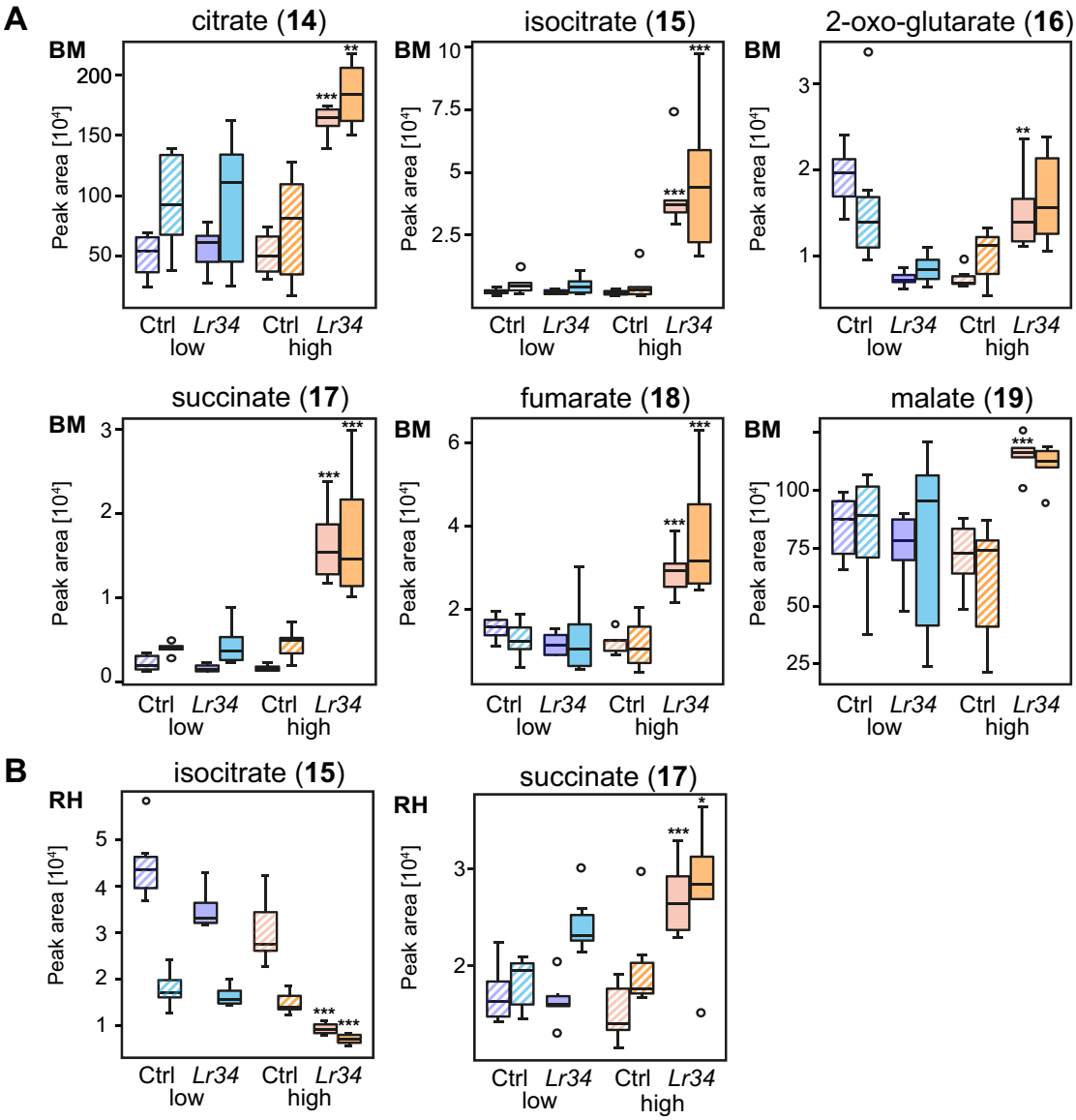
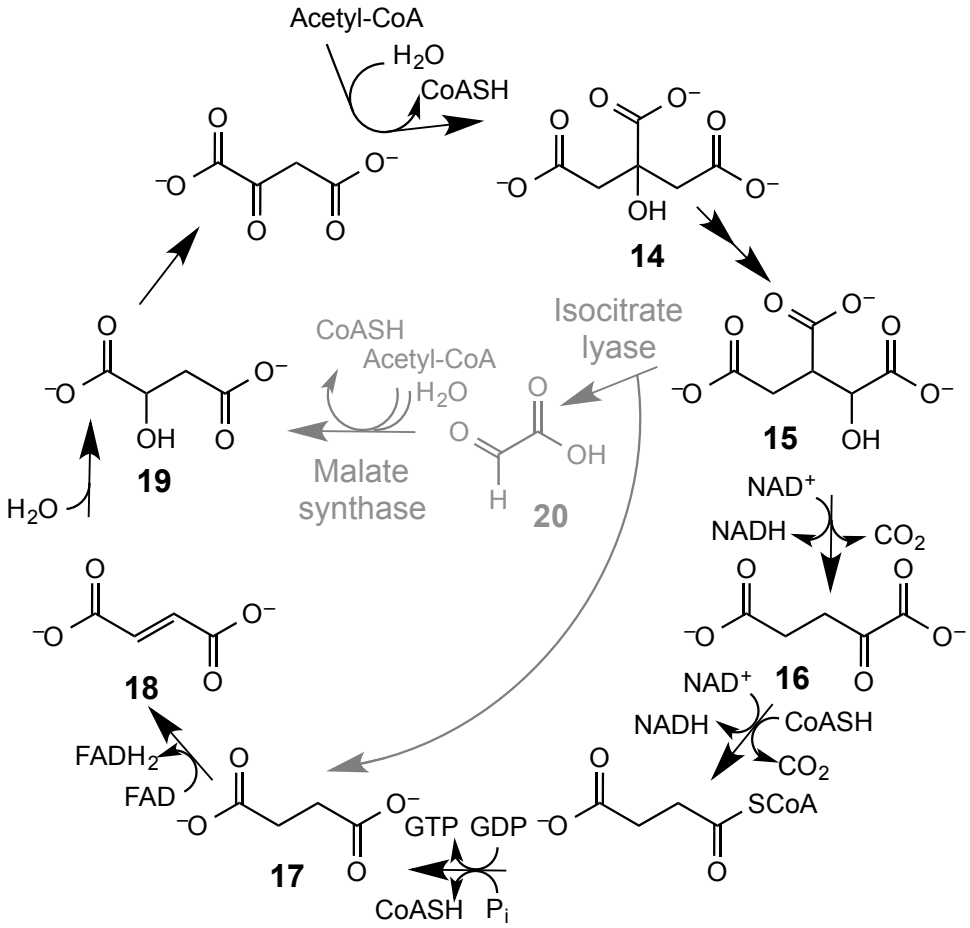


Figure 21: Primary metabolites involved in tricarboxylic cycle are upregulated in high *Lr34* expression levels in mock-infected barley (A), primary metabolites involved in glyoxylate cycle in rice (B).



Scheme 6: Glyoxylate cycle (grey) and tricarboxylic cycle (black); citrate (**14**), isocitrate (**15**), 2-oxo-glutarate (**16**), succinate (**17**), fumarate (**18**), malate (**19**) and glyoxylate (**20**).

The concentration of succinate was increased in barley and rice plants with high *Lr34* expression level grown under different growth conditions — in BH, BM and BI (Supplementary figure 14) and in RH (Figure 21). The increase of succinate (**17**) might be connected to the tricarboxylic acid cycle or to the glyoxylate (**20**) cycle, which are important pathways in the central metabolism of plants (Scheme 6).

Additionally to the succinate (**16**) accumulation, other intermediates of the tricarboxylic cycle, namely citrate (**14**), isocitrate (**15**), 2-oxoglutarate (**16**), succinate (**17**), fumarate (**18**) and malate (**19**), were upregulated in some experiments of high-expressing *Lr34* barley upon mock- or pathogen infection.

Succinate (**17**) was also upregulated in high *Lr34*-expressing rice plants grown under RH condition. In contrary, levels of isocitrate (**15**) were decreased in high expressing *Lr34* rice grown on hydroponics (Figure 21, Scheme 6). In accordance with this observation, transcriptomics experiments performed by collaborators with transgenic *Lr34* rice revealed

overexpression of the isocitrate degrading enzymes isocitrate lyase and malate synthase. These two enzymes are key for the glyoxylate (**20**) cycle that produces succinate (**17**) and malate (**19**), which can be converted to phosphoenolpyruvate to enter the gluconeogenesis that produces glucose [152]. Transcripts of isocitrate lyase and malate synthase were reported to activate gluconeogenesis during plant pathogenesis [153].
Similar effects as in rice were also observed in *Lr34* wheat with increased concentration of succinate and decreased concentration of isocitrate, which is described in 4.7.
Contrasting to rice, a prominent decrease of isocitrate lyase gene expression (log2 –6.8) was observed in *Lr34* barley [27], which matched the observation of increased isocitrate concentration in plants grown under mock-infected condition.

Induction of the tricarboxylic acid cycle and the glycolysis (Supplementary figure 15) in high expressing *Lr34* barley reflects an increase in energy demand. Changes for those two basic pathways were previously reported in different plant-pathogen interactions, amongst them powdery mildew infected barley, suggesting a common response during early infection phase [58]. The role of primary metabolism is to provide cellular energy, for example needed for plant defense response. The energy is used to activate a plethora of infection-related activities, e.g. to express hundreds of genes during defense response [142], to remove carbon and nitrogen compounds from the infection site to starve the pathogen [154], or for cell wall fortification with callose [155, 156], to produce ROS to inhibit the pathogen or to mediate programmed cell death and hypersensitive response [157]. As a consequence, defense responses appear to impose a fitness cost in terms of trade-off in biomass [116].

The disaccharide trehalose was found to be upregulated in high *Lr34* expressing barley upon mock- and pathogen infection.
Oligosaccharides play important roles in pathogen defense. For example, in wheat plants that were partially resistant against powdery mildew, trehalose concentration was increased, similarly to the observation in mock- and pathogen-infected barley containing high expression levels of *Lr34*. These wheat plants were shown to have induced phenylalanine ammonia lyase (PAL) and peroxidase activities [158].
Cell wall invertase enzymes cleave sucrose yielding the monosaccharides glucose and fructose, which are taken up into plant cells by hexose transport proteins. The hexoses are further metabolized intracellularly, e.g. for biosynthesis of secondary metabolites like formation of callose or phenolic compounds, but also act as signaling molecules [159].

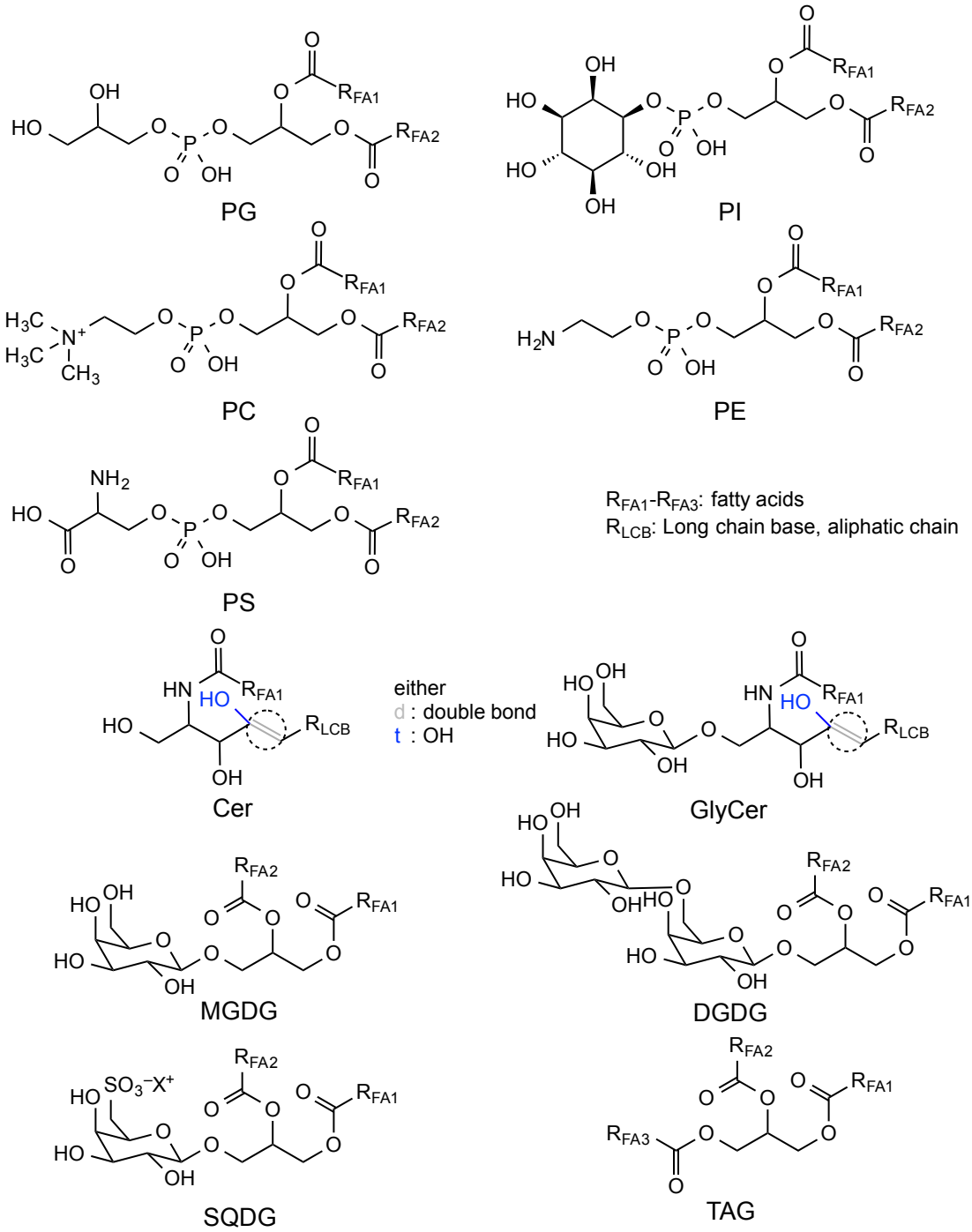
Overexpression of cell wall invertase in rice resulted in constitutive defense-gene activation and higher resistance to rice blast [159].
However, increase of metabolite concentrations of fructose, glucose and threhalose was also observed in a study related to Arabidopsis senescence [160].

In rice, the general response of primary metabolites was modest; few metabolites were regulated such as citrate (**14**), isocitrate (**15**) or succinate (**17**) described above. This was in agreement with a previous study, which investigated rice blast disease in the *Poaceae* plant hosts rice, barley and *Brachypodium distachyon*, and only obtained little discrimination of the metabolic profiles of healthy and infected leaves harvested 24 h after infection. In contrast, large metabolic changes were observed after 72 h, a time point at which the pathogen biomass was increased and the first lesions appeared [40]. So, the moderate metabolic changes observed in this work for rice plants harvested 28 h after rice blast infection corresponded with previous reports.

4.3 Metabolic profiling of lipids

To investigate the metabolic response of lipids in the context of *Lr34* resistance, metabolic profiling of lipids was performed with the apolar extraction fraction by UHPLC-HR-MS. Lipid identification was based on an MPIMP in-house library [144, 161], resulting in quantification of the abundance of 156 lipid species of a variety of lipid classes in barley and rice (Supplementary table 5). General structures of the observed lipid classes are depicted in Scheme 7. The individual lipid species diverged in the fatty acids (FA) chain lengths and their degrees of unsaturation, which are reflected in the notation of the number of carbons before and the number of double bonds after the colon (e.g. PG 32:1 abbreviates a phosphatidyl-glycerol with 32 C-atoms and one double bond). Phospholipids are involved in the lipid bilayer formation of cell membranes, and the major phospholipid classes in plants include phosphatidyl-glycerols (PG), -inositols (PI), -cholines (PC), -cholines having lost one FA chain (lysoPC), -ethanolamines (PE), -ethanolamines having lost one FA chain (lysoPE) and -serines (PS). Further plant lipids, present in plant cell membranes, are sphingolipids derived from ceramide (Cer), glucosyl ceramide (GlyCer), sterol and sterol-esters. Photosynthetic thylakoid membranes are constituted mainly of the galactolipids mono- and di-galactosyl diacylglycerols (MGDG, DGDG) but also contain sulfoquinovosyl diacylglycerols (SQDG) [162]. Additionally, a number of triacylglycerols (TAG), storage lipids, were detected.

Generally in plant metabolism, changes in lipid levels are linked to alterations in plant development and growth, but also to responses of plants to environmental stress such as cold, heat/drought, phosphorus deficiency or salinity [161, 163-165]. The importance of lipids in relation to biotic stresses such as pathogen or herbivore attacks were also highlighted. In biotic stress, lipids were involved in plant defense as components of mechanical barriers, as signaling molecules initiating other defense responses, or as triggers of programmed cell death in hypersensitive response, and even as antifungal or antimicrobial components or as stress mitigators [166].



Scheme 7: General lipid structures of phospholipids PG, PI, PC, PE, PS, sphingolipids Cer, GlyCer, galactolipids MGDG, DGDG, SQDG and TAG.

To explore the data characteristics of the identified lipids, an exploratory data analysis by PCA was performed. A representative PCA result of lipid datasets of barley and rice plants grown under mock-infected conditions is shown in Figure 22. In barley, the corresponding PCA score plots revealed a most distant separation of *Lr34* plants with high expression level from the control and from the *Lr34* plants with low expression level. For the PCA of

barley, the first and the second principal components (PC1 and PC2) explain more than 91% of the variance present in the dataset. The high *Lr34* expression line was clearly separated by PC1 from all other samples. In rice, no clear group separation was obtained, although most of the variance (>62%) of the dataset is captured in PC1 and PC2, indicating weak group specific lipid alterations. Thus, the biggest differences in lipid profiles were expected in high *Lr34* expressing barley.

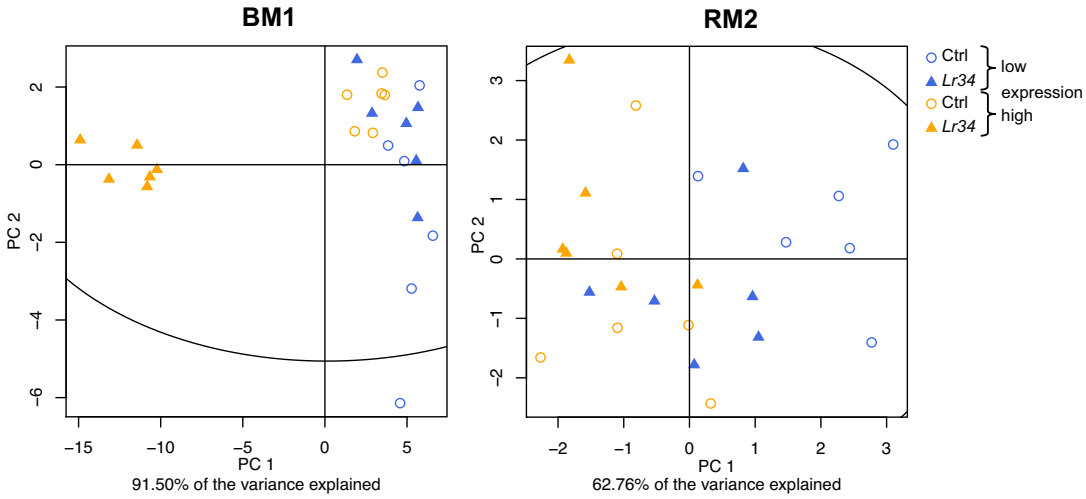


Figure 22: Representative PCA score plots of the identified lipids in mock-infected growth conditions of barley (left) and rice (right). Replicates of the low expression levels are symbolized in blue, replicates of high expression levels in orange, controls as circles and *Lr34* as triangles. A clear group separation was obtained in barley for the high expression level, whereas other groups in barley and rice could not well be distinguished.

To evaluate the changes in lipid levels of plants grown under different growth conditions, the same strategies (A) and (B) were used as described in the previous section for primary metabolites (4.2), and the corresponding results are shown in Figure 23.

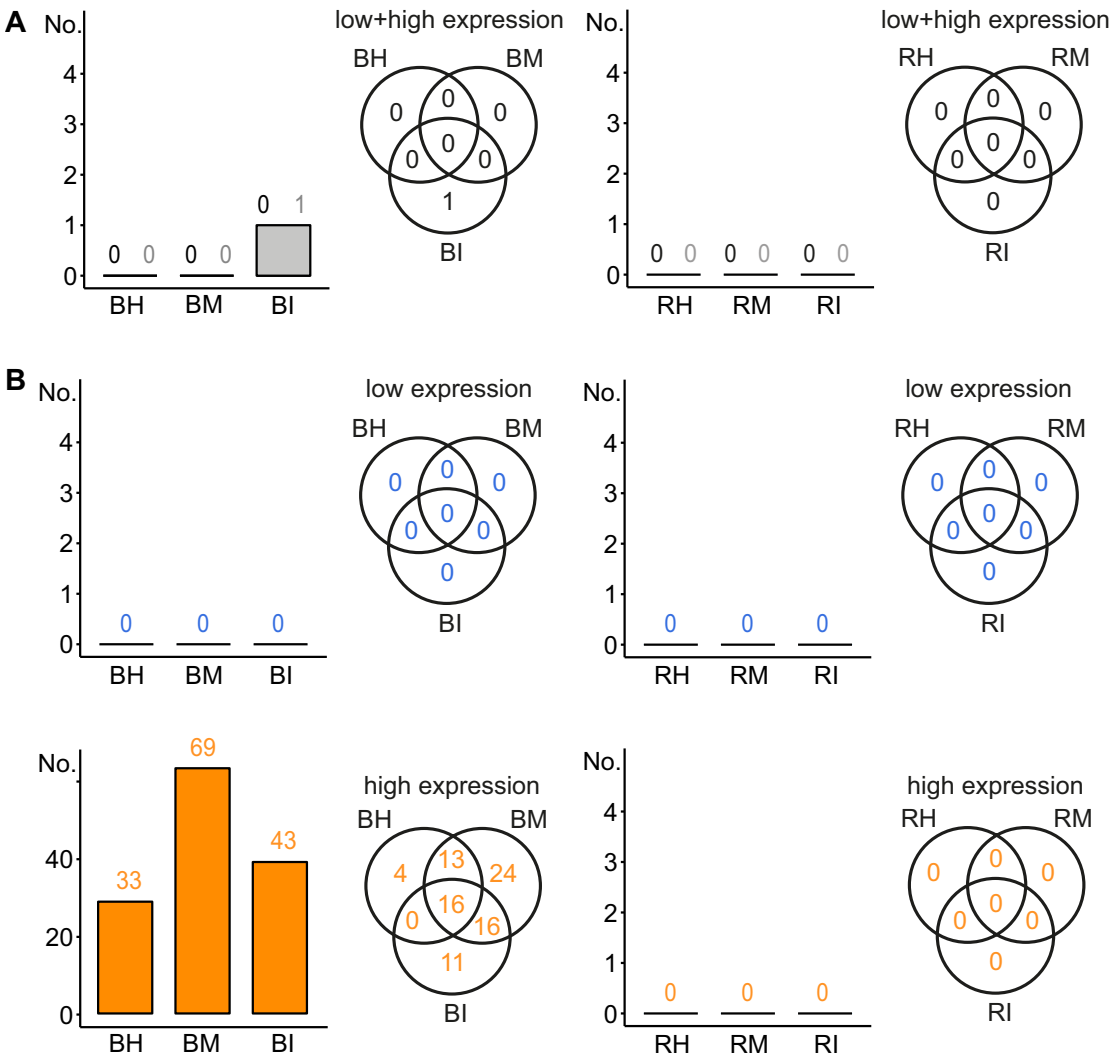


Figure 23: Commonly altered lipids with same regulation tendency and FC > 0.3: based on (A) low and high expression levels of barley and rice with overlaps of 4 and 3 groups (black and grey); (B) separate expression levels (low – blue, high – orange).

As a result of strategy (A), which compares simultaneously both expression levels and both batches of each growth condition, no lipid was found to be consistently up- or downregulated in plants of both low- and high expression levels, neither in barley nor in rice. Evaluation of low- and high-expressing *Lr34* plants according to strategy (B) did not yield any commonly altered lipids for plants with low expression level of *Lr34* — in both cereal species barley and rice. Also no common lipid changes were observed in rice plants of high expression level. In contrast, an overlap of sixteen consistently upregulated lipids was detected for barley plants with high *Lr34* expression level grown under the three conditions BH, BM and BI. The fold changes of the individual lipids are shown in Figure 24, where red and blue colors denote up- and downregulation in *Lr34* plants compared to the respective controls and asterisks indicate the significance. The sixteen consistently

changed lipids were marked with yellow frames and contained each two PCs (34:1, 36:3) and PEs (34:1, 36:3), one Cer (t18:0/c22:0) and eleven TAGs (50:6, 52:5, 52:6, 52:7, 52:8, 54:5, 54:6, 54:7, 54:8, 56:5, 56:8).
Regarding the signal intensities of the identified lipids, the ionization efficiency was reported to be reproducible within a given lipid class comprising the same head group, allowing a direct comparison of the lipid abundances [161]. For the TAGs, the intensities of the triacylglycerol signals are indicated with a bar plot at the right side of Figure 24D. The most intense TAGs, including TAG 52:5, 52:6, 54:6, 54:7, and 54:8, were among the consistent lipid changes for barley plants with high *Lr34* expression.

No clear statement about the biological relevance of the changes in lipid metabolism in context with *Lr34* is possible, but the observed lipid differences indicate connections with plant defense responses and with the strong leaf-tip necrosis phenotype in barley.
Although TAGs serve as efficient high-density carbon and energy sources for reduced carbon and are predominantly found in seeds, pericarps and pollen, they are also found in leaves. In some plants, TAGs are formed as intermediates upon membrane breakdown during stress or senescence. A study reported that TAGs in leaves of various plants differ from senescence-associated TAGs, accumulating during extended periods of stress or at late leaf development. While fatty acid profile of TAGs observed in senescent *Arabidopsis thaliana* leaves matched the fatty acid composition of the galactolipids in thylakoids [167], the chemical composition of the TAGs present at early leaf development was different compared to the composition at later developmental stages, containing shorter and more saturated FAs [168]. TAGs may represent an intermediate within the mobilization of membrane lipid carbon via de-esterification of galactolipids and subsequent conversion of the free fatty acids to sucrose within the glyoxylate cycle in the peroxisome [167].
The increased concentrations of TAGs found in this experiment contained either 50, 52, 54, or 56 carbons in fatty acid chains, indicating combinations of fatty acids with chain lengths of 16-C, 18-C, and 20-C in ratios of 2:1:0, 1:2:0, 2:0:1, 1:1:1, 0:3:0, and 0:2:1, respectively. Previously reported TAGs in senescent *Arabidopsis* contained increased levels of 16:0, 18:1, 18:2 and 18:3 fatty acids. Thus, the increase of TAG levels in high *Lr34*-expressing barley was assumed to be associated with the leaf senescence phenotype and the increased energy demand needed in connection with constitutively activated plant defense.

Sphingolipids essentially regulate cell death and pathogen-induced hypersensitive responses. Elevated levels of ceramides were associated in a rather complex relationship with salicylic acid-dependent signaling cascades resulting in enhanced cell death [169]. For example, *A. thaliana* with modified sphingolipid content were recently reported to be affected in tolerance of hemi-biotrophic and necrotrophic pathogens [170]. Hence, the upregulated Cer t18:0/c22:0 might also be involved in high-expressing *Lr34* barley leaf senescence phenotype.

With the method used in this investigation, a range of different lipid classes was detected, but other lipid classes with relation to plant defense — for example phosphatidic acid derivatives — were not observed. Among the pathogen responsive lipids, phosphatidic acid derivatives play a certain role, being both an intermediate metabolite of the biosynthesis of all glycerolipids and a signaling molecule. Phosphatidic acid derivatives are involved in basal defense and non-host resistance against barley powdery mildew infection in *A. thaliana* [171], but also in the regulation of hormone-mediated responses by salicylic- and abscisic acid, and in abiotic stresses [169]. However, the extraction method used in the current work based on MTBE was not well suited for the detection of phosphatidic acid species. Other lipid analyses with an extraction protocol based on a chloroform-methanol mixture, acidified with 0.1 M HCl, showed superior sensitivity for this specific lipid class [172].

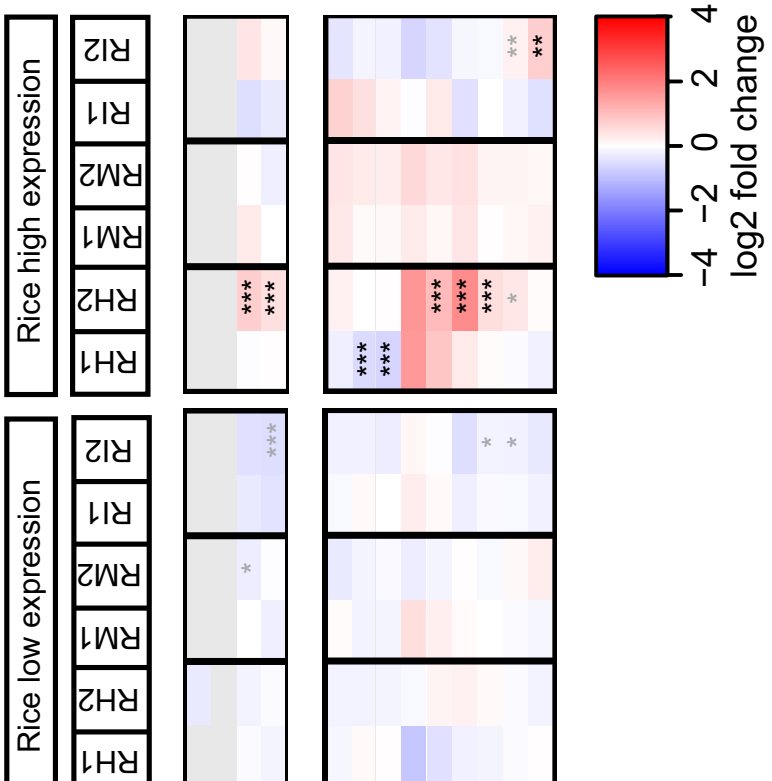


Figure 24: Heatmap of lipid changes in barley and rice in different experiments, with commonly changed lipids highlighted with yellow frame. Lipids are grouped in Cer and PE (A); PC, lysoPC and lysoPE (B); MDGD, DGDG, and SQDG (C); and TAGs (D). GlyCer, PI and PS were only detected in wheat. Ratios of log2-transformed fold changes are given by shade of red or blue according to the scale bar. Data represent mean values of six replicates. Statistical analysis was performed with ANOVA and post-hoc Tukey's test, of which * *p*-value <0.05, ** < 0.01 and *** < 0.001; grey asterisks indicate either fold change below 0.3 or inconsistency of regulation tendency.

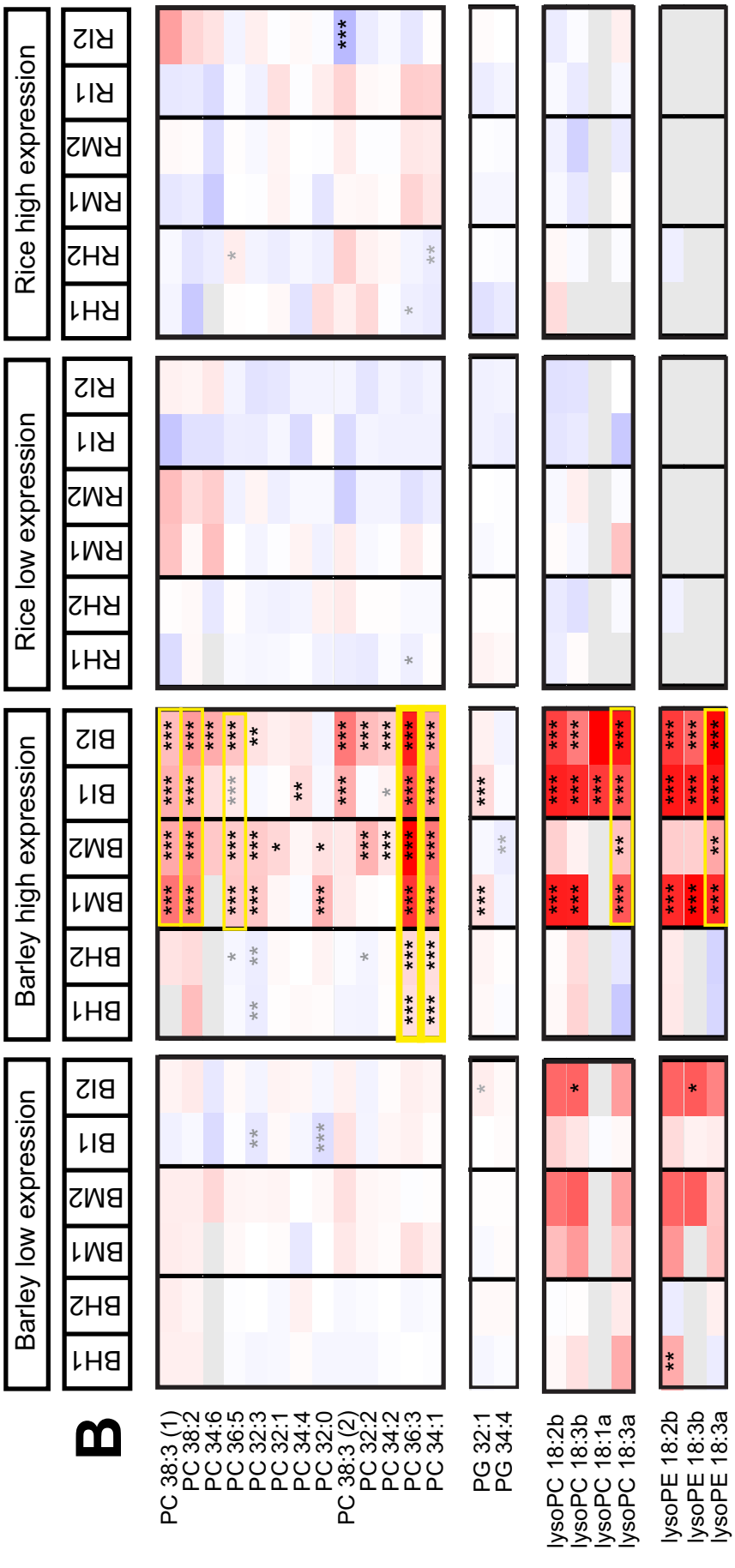


Figure 24, *continued*

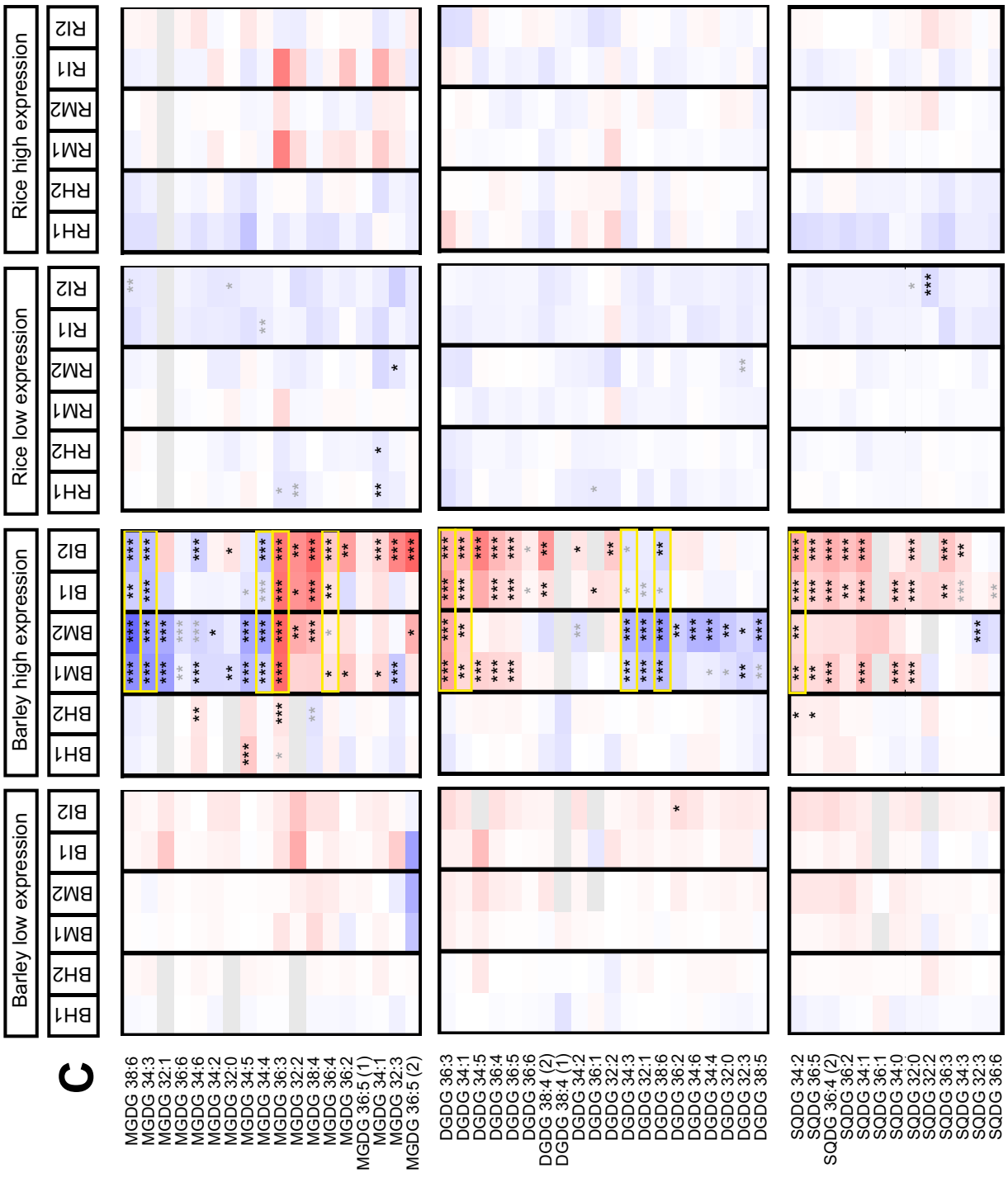


Figure 24, *continued*

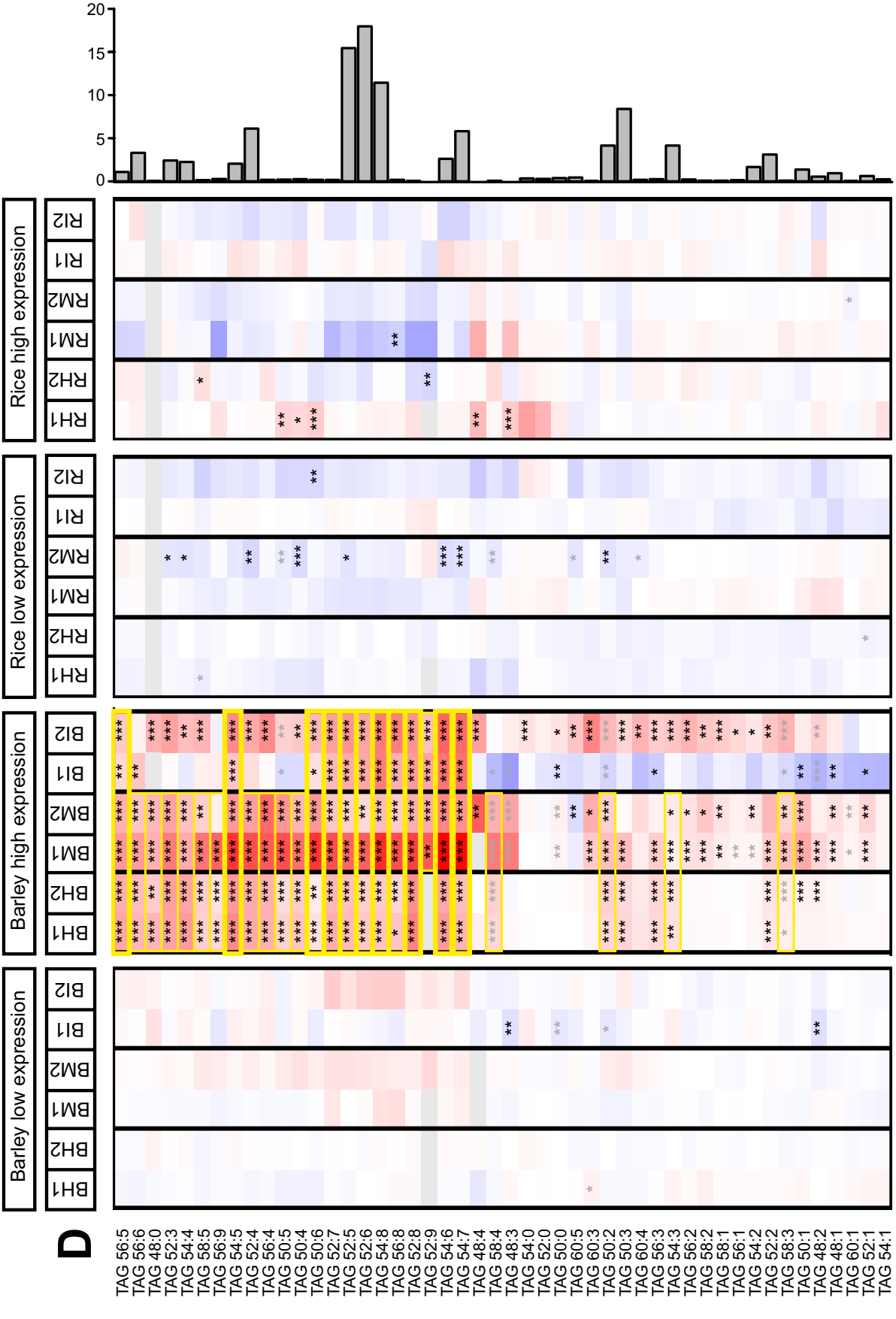


Figure 24, *continued*, the bar plot right represents the relative abundance of each TAG within the TAG lipid class in BM1 Ctrl low expression.

4.4 Metabolic fingerprinting of secondary metabolites

For the investigation of secondary metabolite changes in *Lr34*, a non-targeted metabolic fingerprinting was performed for barley and rice based on a highly reproducible and sensitive UHPLC-method [173]. After preprocessing of the LC-MS data, large tables with a high number of features were obtained that were evaluated by statistical analysis to find features with significant difference between control and *Lr34* groups.

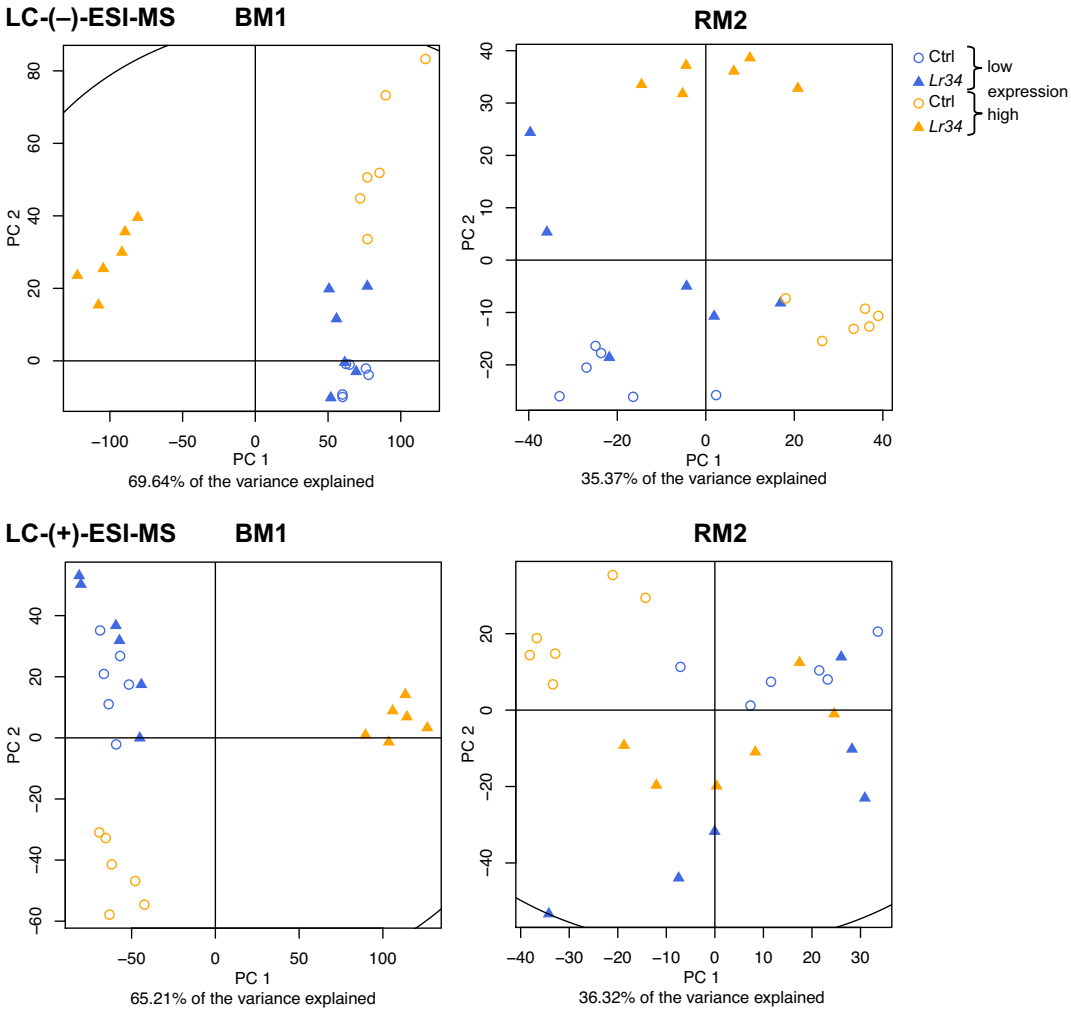


Figure 25: Representative PCA score plots of features of secondary metabolites of barley (left) and rice (right) grown under mock-infected growth conditions, with LC-(–)-ESI-MS shown above and LC-(+)-ESI-MS shown below. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles and *Lr34* as triangles.

An explorative PCA was performed to visualize the metabolic fingerprints of secondary metabolites, and representative score plots are shown in Figure 25. In mock-infected barley, the orange symbols were well separated by PCA, corresponding to control (circles) and *Lr34* (triangles) of high expression levels. The blue motifs that symbolized low

expression levels of control and *Lr34* were not separated. Again, the second principal component (PC2) separates mainly control groups of high and low expression lines. These observations are in line with the results obtained by PCA of the lipid and primary metabolite datasets. In rice, the group separation was not as clear as in barley, but the high *Lr34* expression level showed a trend of separation from the other groups.

Then, the data was investigated by univariate statistical analysis to disclose significantly different secondary metabolite features in each batch. The differences were compared amongst the three growth conditions in barley and rice in analogy to the procedure described for primary metabolites in 4.2.

A single metabolite is often detected as several features caused by in-source fragments or adducts in either both LC-(+)-ESI-MS and LC-(–)-ESI-MS ionization modes or a single ionization mode, depending on the chemical properties of the metabolite. For this reason, significantly changed features were evaluated by comparing data of LC-(+)-ESI-MS and LC-(–)-ESI-MS separately (Figure 26 and Figure 27).

The simultaneous evaluation of the low- and high expressing genotypes according strategy (A) for barley under all three growth conditions including hydroponics, mock– and pathogen infected growth revealed an intersection of a single LC-(–)-ESI-MS feature with *m/z* 504.6033 at RT 0.62 min. This feature eluted in the dead volume of the column, which complicated further annotation.

For rice, no intersection was obtained between the three growth conditions RH, RM and RI in the simultaneous comparison of the low- and high expressing genotypes for both (+)-ESI and (–)-ESI ionization modes. Also the comparison of the low-expressing genotypes according strategy (B) of both barley and rice did not produce an overlap in both ionization modes.

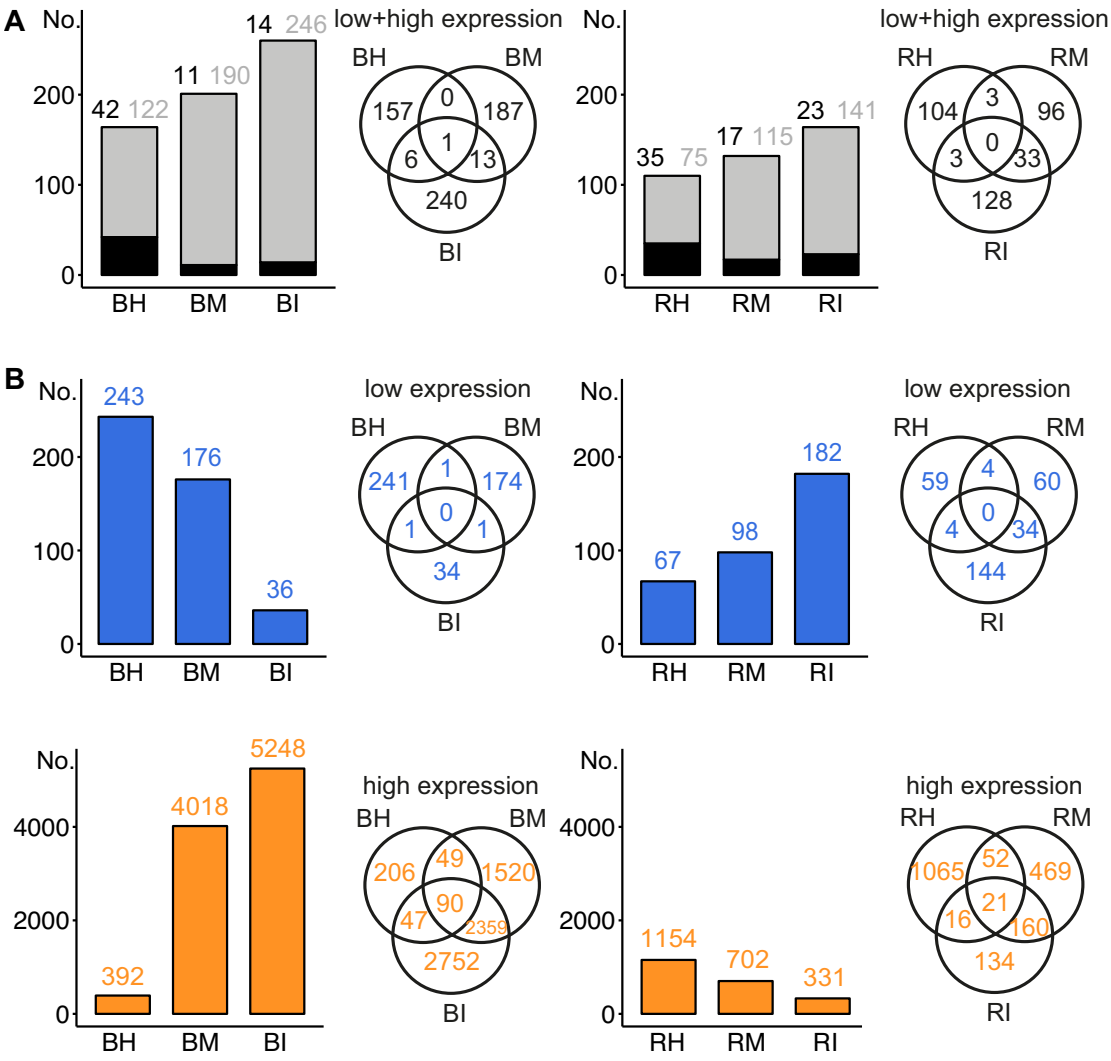


Figure 26: Commonly altered LC-(–)-ESI-MS features of secondary metabolites with same regulation tendency: based on (A) combined low and high expression levels of barley and rice with overlaps of 4 and 3 groups (black and grey) and (B) separate expression levels (low – blue, high – orange).

In contrast, the comparison of the high expressing genotypes produced intersections of features that were changed under all three growth conditions. In the LC-(–)-ESI-MS dataset, 90 features were robustly regulated in barley, and 21 features were robustly changed in rice. In the LC-(+)-ESI-MS data, intersections contained slightly lower numbers of features for barley with 29, and rice with 7 robustly regulated features.

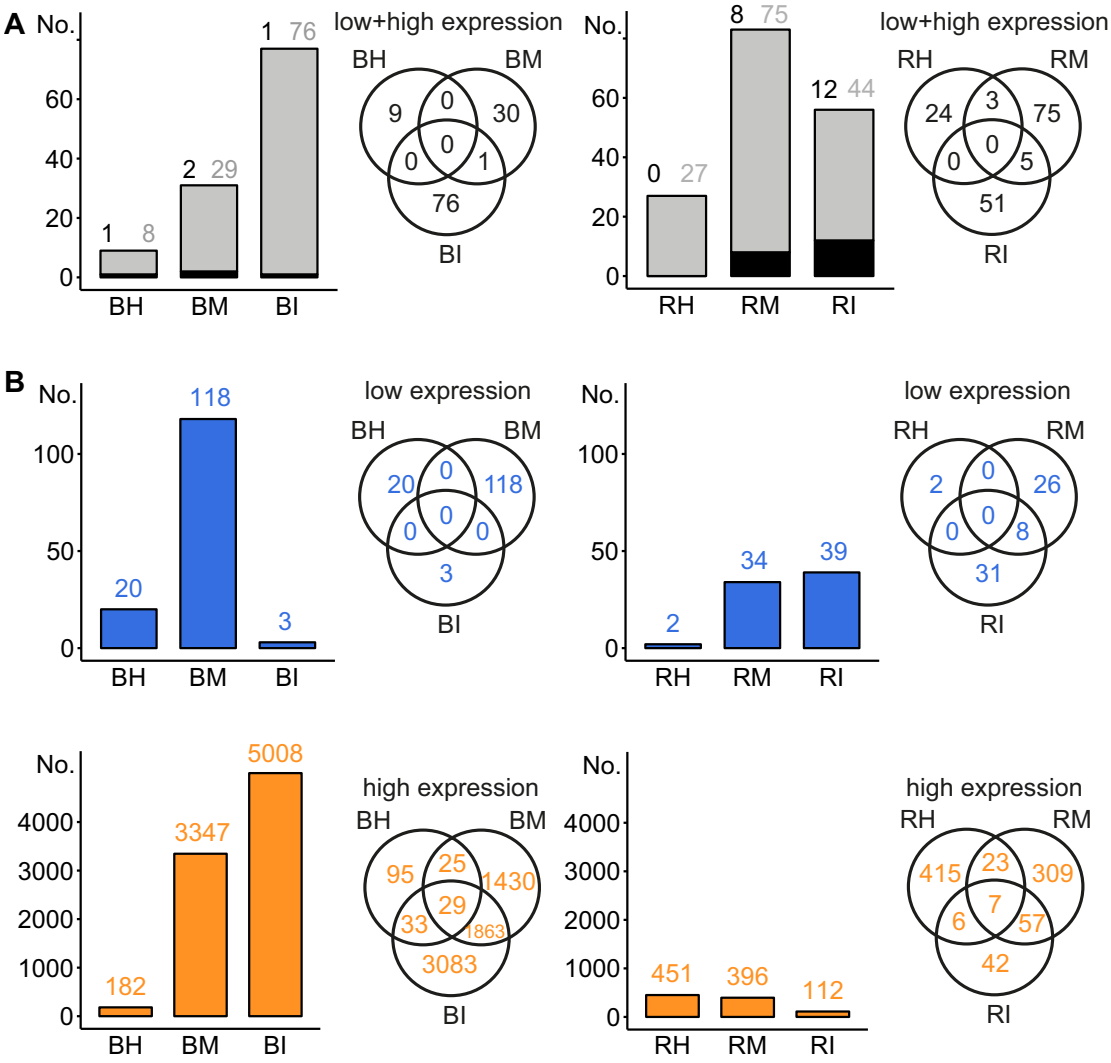


Figure 27: Commonly altered LC-(+)-ESI-MS features of secondary metabolites with same regulation tendency: based (A) on combined low and high expression levels of barley and rice with overlaps of 4 and 3 groups (black and grey) and (B) on separate expression levels (low – blue, high – orange).

Features that were found in the intersections of the high-expressing genotypes of barley and rice were considered as robust metabolic changes appearing in plants grown under all three growth conditions. In order to interpret these differences on a biological level, the aim was to identify the underlying metabolites. However, metabolite annotation of secondary plant metabolites is still a bottleneck of metabolomics. The main reasons for the difficulties in identification are the high chemical diversity of secondary metabolites (estimated 200'000 in the plant kingdom), the low number of available reference substances, and the often low concentration of metabolites. From a mass spectrometry perspective, ESI-MS/MS spectra strongly depend on the acquisition parameters of the instrument, and comparability of recorded spectra with spectra in public databases is limited. Often,

already the evaluation of the molecular mass of a component, which is used in order to calculate a chemical formula, is a difficult task due to in-source fragments or adducts. Furthermore, reactivity of the metabolites may lead to degradation of features of interest, therefore not allowing the isolation of interesting precursor ions for MS/MS fragmentation analysis. As a consequence, a large proportion of the commonly changed metabolites remains still unknown.

The fold change of the LC-(+)-ESI-MS and LC-(–)-ESI-MS features obtained from the intersections of all three growth conditions in high-expressing genotypes of barley and rice were plotted in a heatmap shown in Figure 28. The features are sorted according to RT, and clustered into 67 components according to their retention time and their tendency of common fold change in various conditions. Out of these clusters, 29 components contained more than two features and were thus considered more robust. Two interesting components were selected for closer investigation: component **o** at 3.65 min that was upregulated in high-expressing barley and component **p** at retention time 2.39 min that was downregulated in high-expressing rice (Figure 28, Supplementary table 6).

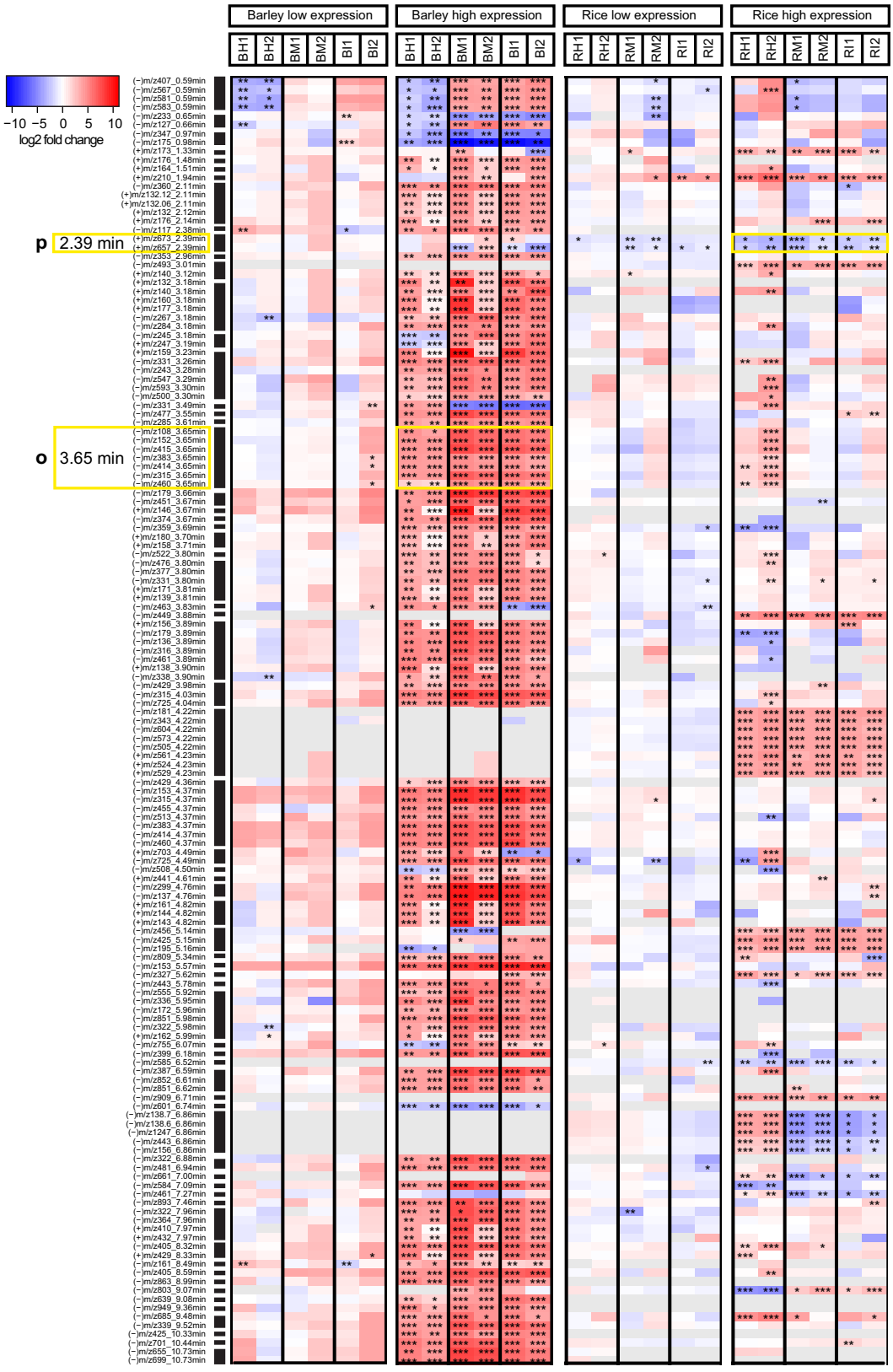
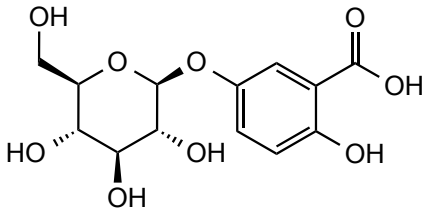


Figure 28: Heatmap of overlapping secondary metabolite features in barley and rice in different experiments. Black bars on the left indicate components. Depicted values are log2-transformed mean fold changes colored according to the scale bar. Significance according Tukey's post-hoc test (* *p*-value <0.05, ** < 0.01 and *** < 0.001).

Based on HR-MS data, component **o** was assigned the chemical formula $C_{13}H_{16}O_9$. Additionally to the $[M–H]^-$ ion at *m/z* 315.0721, a fragment at *m/z* 108.0202 was observed in the ESI-MS data. A database query with the chemical formula $C_{13}H_{16}O_9$ in the *SciFinder* database suggested gentisic acid *O*-glucoside as a possible metabolite. In recently reported mass spectra of gentisic acid 2-*O*- and 5-*O*-*β*-glucosides (**21**) from sugarcane leaves, the fragment at *m/z* 108 was also observed [174]. Thus, component **o** was putatively assigned as gentisic acid *O*-glucoside. In literature, the abundance of the fragment at *m/z* 108 was higher for the 5-*O*-*β*-glucoside than for the 2-*O*-substituted analog. Gentisic acid is the oxidation product of the plant hormone salicylic acid and has been found in its glycosylated form in many plant species such as buckwheat [175].



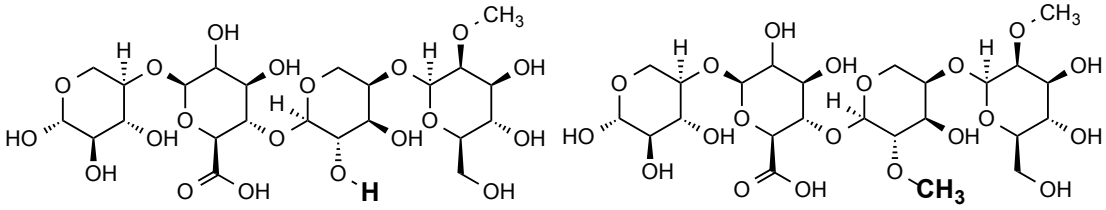
21

Scheme 8: Gentisic acid 5-*O*-*β*-D-glucoside (**21**)

For component **p** at RT 2.39 min, the chemical formula $C_{23}H_{38}O_{20}$ was calculated based on the HR-MS data of the $[M-H]^-$ ion at *m/z* 633.1890 and of the $[M+H]^+$ ion at *m/z* 635.1863. The *SciFinder* hit list obtained for this formula suggested only four possible structures, including two tetrasaccharides. Additionally to component **p**, a structurally related component **q** ($C_{24}H_{40}O_{20}$) revealed in positive and negative MS/MS numerous identical fragments as well as some that were shifted by 14.0159 Da, corresponding to a CH_2 group (Supplementary figure 17). A *SciFinder* database query suggested several methylated and cyclic tetrasaccharides. For the lichen *Collema flaccidum*, the two tetrasacharides **31** and **32** were described, which would correspond to the chemical formulas of components **p** and **q** [176]. No reference was found for the MS/MS spectra. For both components **p** and **q**, the features that corresponded to $[M+H]^+$ and $[M–H]^-$ were isolated for CID-MS/MS fragmentation. The MS/MS spectra obtained for both components showed a number of highly similar fragments. This inferred a common structural moiety that would correspond to the chemical formula $C_{10}H_{20}O_9$. The remaining structural parts of the molecules would then correspond to $C_{13}H_{18}O_{11}$ for component **p** and to $C_{14}H_{20}O_{11}$ for

component **q**. These remaining structural parts were shifted by CH_2, for example caused by a methoxyl group instead of a hydroxyl group.

From a biological perspective, oligosaccharides may arise from simultaneous hydrolysis of multiple sugar-containing secondary metabolites. They are effective signaling molecules and may trigger strong plant defense responses [116]. For example, a glucopentose purified from enzymatic digest of *β*-glucan from the rice blast fungus induced phytoalexin biosynthesis in rice cell culture [177].

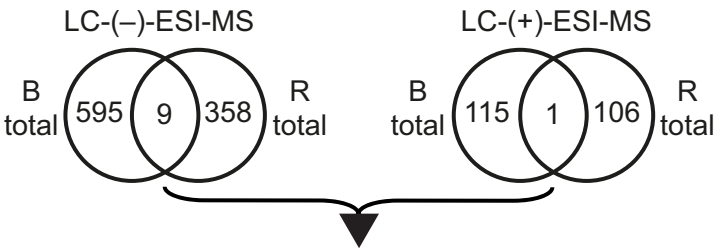


22 **23**

Scheme 9: Structures suggested for chemical formulas of components **p** $C_{23}H_{38}O_{20}$ *O*-2-*O*-methyl-*β*-D-mannopyranosyl-(1→4)-*O*-*α*-D-arabinopyranosyl-(1→4)-*O*-*β*-D-glucopyranuronosyl-(1→4)-D-xylopyranose (**22**) and **q**, $C_{24}H_{40}O_{20}$: *O*-2-*O*-methyl-*β*-D-mannopyranosyl-(1→4)-*O*-2-*O*-methyl-*α*-D-arabinopyranosyl-(1→4)-*O*-*β*-D-glucopyranuronosyl-(1→4)-*β*-D-xylopyranose (**23**) [176].

The current dataset presents an opportunity to compare the metabolic response of *Lr34* in barley and rice. All features that were found different between control and *Lr34* plants in barley and rice in any of the conditions were summed up based on Figure 26A and Figure 27A. For the LC-(–)-ESI-MS data, these features were summed up to 604 and 367 for barley and rice, respectively, and had an intersection of 9 features. The LC-(+)-ESI-MS data were added up to 161 and 107 features for barley and rice and had an intersection of a single feature.

One of the features found in both barley and rice was detected in LC-(–)-ESI-MS with *m/z* 771 at RT 5.22 min, and is referred to as component **r**.



	ESI	PeakID	*m/z*	RT [min]	
	(+)	Peak_027686	208.8504	0.61	dead volume
	(–)	Peak_055876	346.7777	0.61	
	(–)	Peak_055357	344.7813	0.61	
	(–)	Peak_017733	172.8884	0.61	
	(–)	Peak_017402	170.8913	0.61	
	(–)	Peak_063147	377.0852	0.81	
	(–)	Peak_038275	230.0481	2.45	
r	(–)	Peak_155329	771.1997	5.22	
	(–)	Peak_188760	981.3542	6.21	
	(–)	Peak_188731	981.2499	6.24	

Figure 29: Comparison of secondary metabolite features detected in simultaneous consideration of low and high expression levels for barley and rice, based on **Figure 26**A and **Figure 27**A.

Besides the component **r**, a group of related secondary metabolite components **s–ah** were investigated and found to be significantly downregulated in *Lr34* compared to control groups in different experiments of barley and rice, mainly in the high-expressing genotypes (Figure 30).

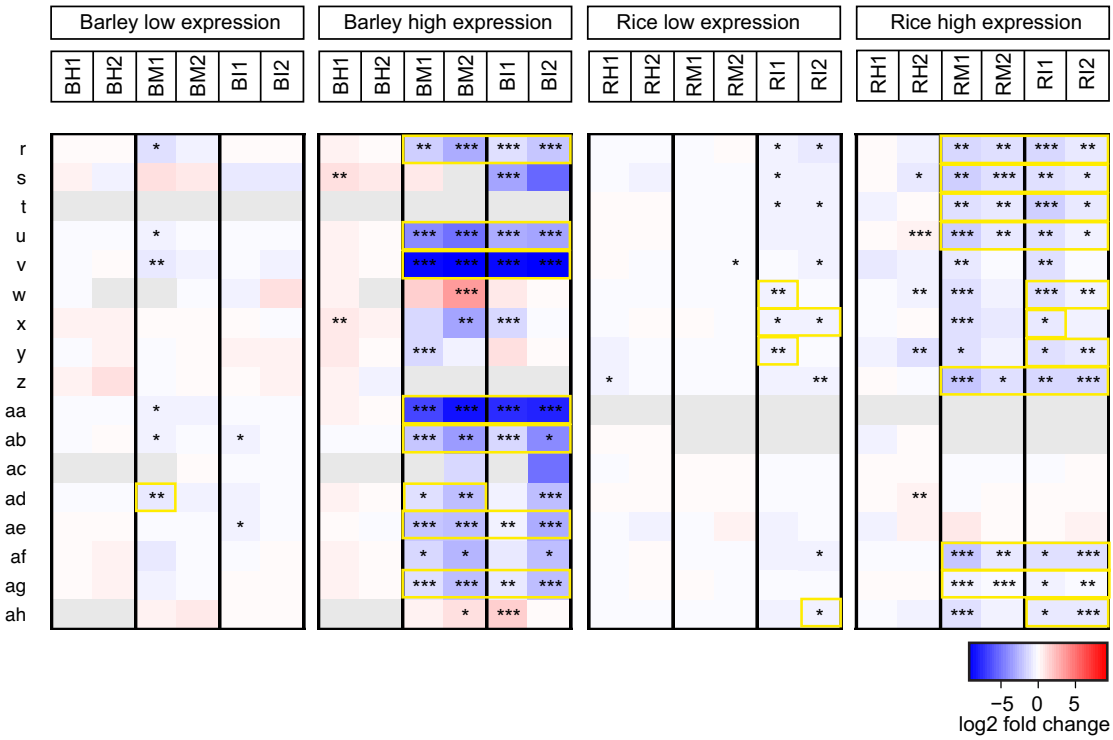
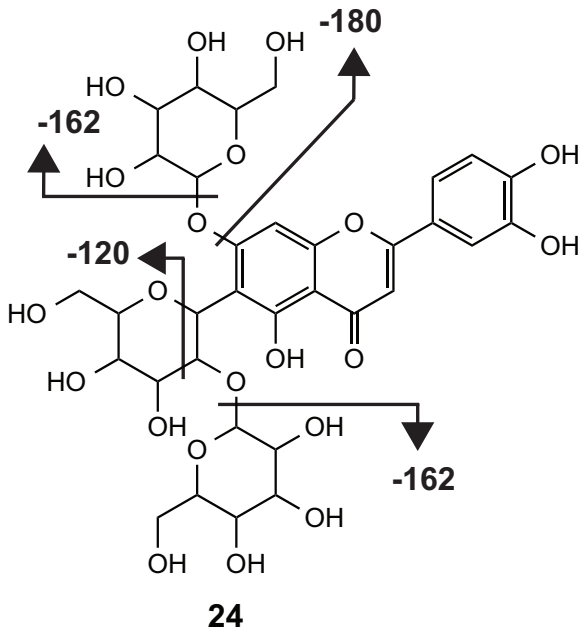


Figure 30: Heatmap of secondary metabolite components **r**-**ah** in barley and rice in different experiments, with commonly changed lipids highlighted with yellow frame. Ratios of log2-transformed fold changes are given by shade of red or blue according to the scale bar. Data represent mean values of six replicates. Statistical analysis was performed with ANOVA and post-hoc Tukey's test, of which * *p*–value <0.05, ** < 0.01 and *** < 0.001.

In a further step, the chemical formulas were calculated for components **r**-**ah** based on HR-MS data, and MS/MS spectra were recorded in order to annotate secondary metabolite identities. A list of features corresponding to components **r–ah** can be found in Table 12, and detailed MS/MS spectra are displayed in Supplementary figure 18 and following.



Scheme 10: Fragmentation of component **r** ($C_{23}H_{38}O_{20}$) putatively assigned as isoorientin-7,2”-di-*O*-glucoside (**24**).

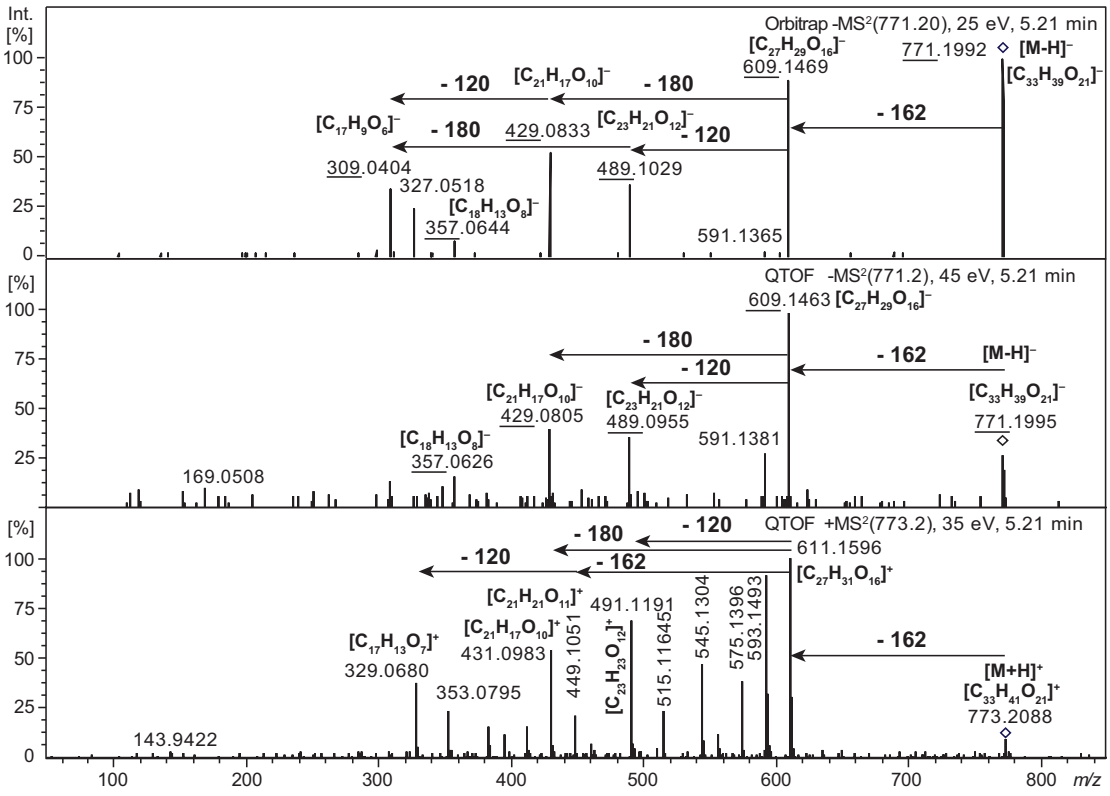
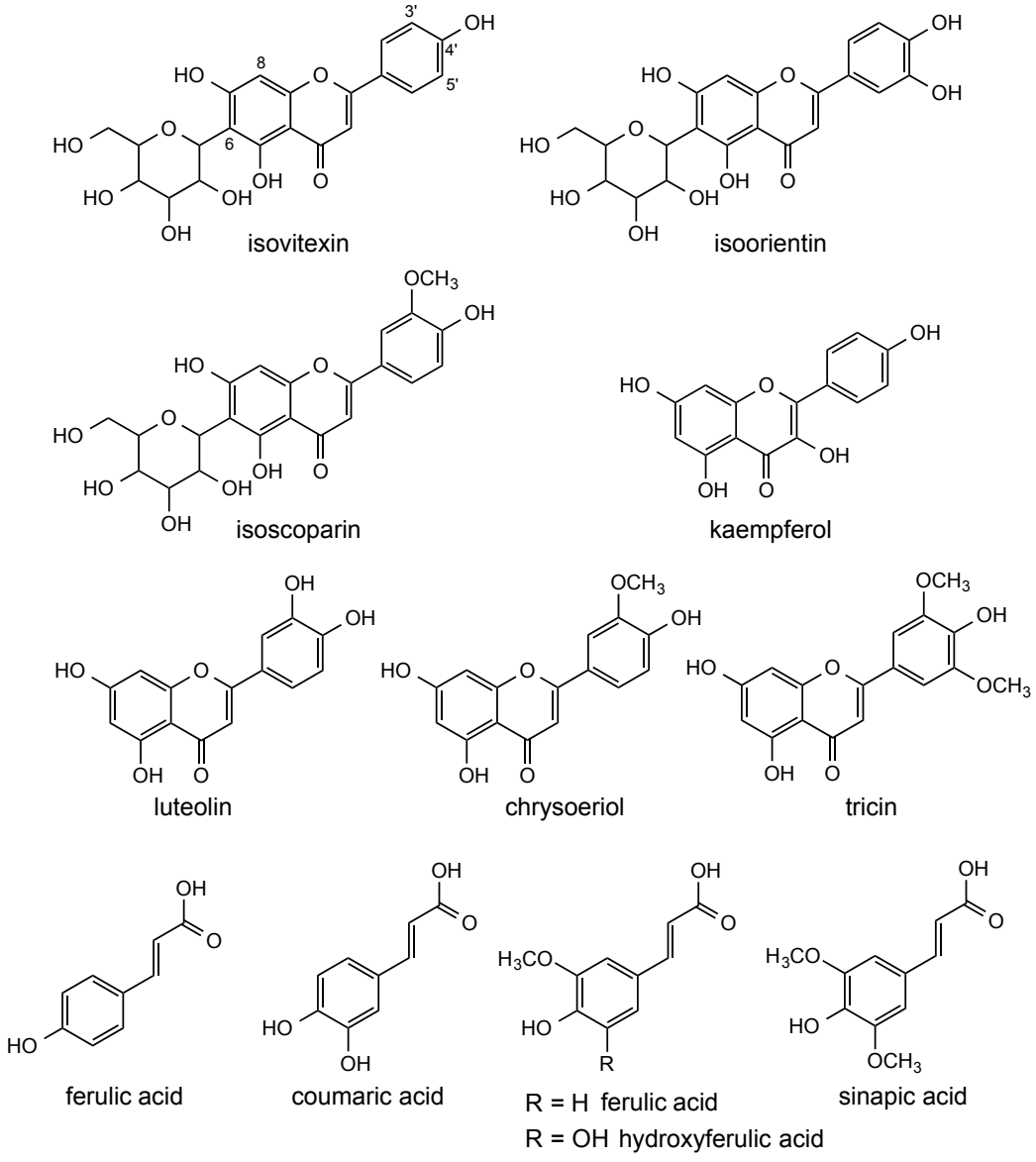


Figure 31: MS/MS spectra of component **q** $C_{23}H_{38}O_{20}$ putatively assigned as isoorientin-7,2”-di-*O*-glucoside (**24**); MS/MS from Orbitrap (top, $[M-H]^-$) and QTOF (middle $[M-H]^-$, bottom $[M+H]^+$). Fragments reported in literature are underlined [74].

Components **r–ah** were putatively assigned as flavonoids based on MS/MS spectra similarities of the respective fragmentation patterns with literature. More specifically,

flavonoids **r–y** and **aa–ae** showed characteristic fragmentation behavior of *C*–glycosylated flavones.

For example, component **r** showed upon (+)- and (–)-ESI-MS/MS fragmentation of $[M+H]^+$ and $[M–H]^–$ characteristic losses of –120 Da that would correspond to the rupture of the *C*-glycosidic sugar (Scheme 10, Figure 31). If the (–)-ESI-MS/MS spectrum of $[M–H]^–$ of component **r** was recorded on an Orbitrap instrument, it matched well with a reference spectrum of isoorientin-7,2”-di-*O*-glucoside (**24**) reported in barley leaf extracts in literature [74].



Scheme 11: Suggested flavone core structures and suggested acylations.

Putative annotation of components **r**–**ag** relied on mass spectra from literature, thus corresponding to metabolite identification level 2 (chapter 1, 2.6).

Besides the flavone core structure isoorientin described for component **r**, reference spectra of the putatively annotated components **r**–**y** and **aa–ae** also suggested the core

structures isovitexin and isoscoparin (Scheme 11). All three flavones contain a *C*-glucoside on position 6, but the exact glycosylation position of components **r**–**y** and **aa–ae** could not be determined based on the available data.

Components **t**–**w**, **y** and **aa–ae** were found to be flavones that are substituted at the glycoside with acylations of hydroxycinnamic acids coumaroyl, caffeoyl, feruloyl, hydroxyferuloyl or sinapoyl, with the corresponding acyl-ions detected in (+)-ESI-MS/MS. In literature, isovitexin-2”-*O*-glucoside, isovitexin-7-*O*-glucoside and isoorentin-7-*O*-glucoside were mentioned as the dominant flavonoids in barley [74]. Thus, component **x** was assumed to be an isovitexin-*O*-glucoside. In the present study, metabolite annotation was based on MS/MS experiments, which only provided limited information about the exact connectivities of the metabolite. In contrast, MS^n fragmentation spectra allow gaining further structural information about glycosidic linkage and substitution, but often they are not sufficient to attain complete certainty about the connectivities that is only achieved by NMR analyses of purified compounds. By NMR spectroscopy, the presence of the flavones luteolin, apigenin and chrysoeriol was confirmed in barley leaves [74]. Regarding the acylation position, 7-*O*-[6”-acyl]-glucosides and 7-*O*-[6”-acyl]-glucosyl-4′-*O*-glycosides [74] had been reported in barley that were characterized by NMR and mass spectrometry, and interglycosidic bonds in barley are known to form as glycosyl(1→6)glycosides or glycosyl(1→2)glycosides [178]. Thus, the flavones observed in the current study may be modified at acylation positions and with interglycosidic linkages described in literature, but the assignment remains speculative.

The biological interpretation of the decrease of flavones observed in transgenic *Lr34* barley and rice is not straightforward. Because flavonoid biosynthesis was generally enhanced at the transcript level and flavonoids are known to be defense compounds against pathogens, an enhanced level of flavonoids was expected.

More specifically, the gene expression of phenylalanine ammonia lyase (PAL) was found strongly upregulated in *Lr34* barley seedlings compared to a sister line [27]. PAL is a key enzyme at the beginning of the phenylpropanoid pathway, and further genes related to the flavonoid pathway were differentially expressed. Accordingly, an increase rather than a decrease in flavonoid concentration was expected in *Lr34* plants.

Generally, flavonoids are involved in plant defense against various abiotic and biotic stresses including UV radiation and pathogens. Flavonoids function mainly as anti-oxidants and scavengers of free radicals, including ROS. Several flavonoids including naringenin, isorhamnetin, kaempferol and methoxyflavone derivatives were associated to

barley resistance to fungal infection with *Fusarium* [66]. On the other hand, anti-pathogenic activities of flavonoids were shown to depend on their structure, and strongest anti-fungal effects were found for unsubstituted flavones and flavanones [179, 180]. Also aglycones of apigenin and luteolin were shown to be pathogen-inducible antifungal compounds and considered as sorghum phytoalexins [181]. Thus, the decrease of *C*-glycosylated flavones might be linked to antifungal activity of their degradation products like their aglycones, which were not yet identified in this dataset. For example, the downregulated flavones could undergo either degylcoslyation, hydroxycinnamic acid derivatives cleavage, further glycosylation with additional sugars, or modifications by methylations, hydroxycinnamic acid derivatives etc.

Flavonoid content further depends on the developmental stage of the plants and young leaves of *Cistus ladanifer* showed higher flavonoid synthesis [182]. Under elevated CO_2 concentration, wheat flavonoid contents of isoorientin and tricin decreased in mature leaves by one third compared to flag leaves and isoorientin almost disappeared in senescing leaves [183]. Flavonoid concentrations are known to fluctuate in response to plant stress [160]. Thus, flavone decrease might be associated with the leaf senescence phenotype in high-expressing barley. We also observed a similar tendency of decline in flavone content for high-expressing rice, even though the harvested leaves were not senescent.

It is important to note that the putatively identified, downregulated components **r**–**ah** only represent a small number of differentially regulated features detected in the secondary metabolite dataset. Because still a large number of components remain unknown, the biological interpretation of the current results cannot be done conclusively. The unknown metabolites might reveal additional hints for explicit explanations of the currently observed flavone decline, and further expose pathways associated with *Lr34*-mediated resistance.

Table 12: Component list of putatively annotated secondary metabolites in barley and rice sorted according retention times. References Piasecka [74], Dong [85], Yang [184], Gong [185], Matsuda [84].

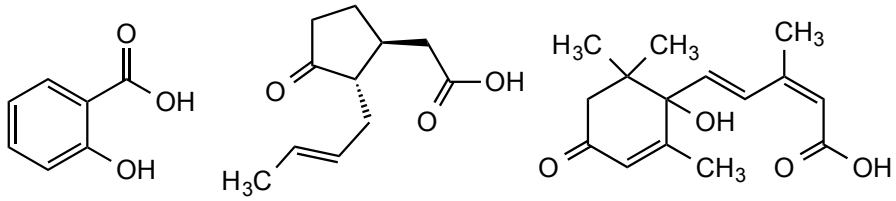
Component	Formula	Ion	Peak *m/z*	Deviation *m/z* [ppm]	Peak RT [min]	Intensity (BH low) Ctrl	*p*-value	FC	B	R	Putative annotation	Annotation level	Reference
r	$C_{33}H_{40}O_{21}$	$[M-H]^-$	771.19971	1.0	5.22	5589	5.1E-03	0.33	X	X	Isoorietin-7,2"-di-O-glucoside	2	Piasecka
		$[M+H]^+$	773.21381	0.4	5.21	12682	7.4E-04	0.20					
		$[M+Na]^+$	795.19245	-3.8	5.21	13309	NA	NA					
s	$C_{33}H_{40}O_{21}$	$[M-H]^-$	771.19976	1.1	5.56	872	6.7E-03	0.44		X	Isoorientin-di-O-hexoside	3	–
		$[M+H]^+$	773.21517	2.2	5.56	2884	1.6E-03	0.40					
		$[M+Na]^+$	795.19683	1.8	5.55	NA	NA	NA					
t	$C_{33}H_{40}O_{21}$	$[M-H]^-$	771.19986	1.2	5.78	NA	6.7E-06	0.34		X	Isoorientin-7-*O*-diglucoside	2	Piasecka
		$[M+H]^+$	773.21361	0.2	5.79	NA	5.0E-05	0.32					
u	$C_{37}H_{38}O_{19}$	$[M-H]^-$	785.20072	9.3	6.74	17101	5.5E-09	0.05	X	X	Isovitexin 7-*O*-[6″-hydroxyferuloyl]-glucoside	2	Piasecka
		$[M+H]^+$	787.21272	6.0	6.74	22660	2.9E-03	0.04					
		$[M+Na]^+$	809.19038	0.5	6.74	3088	NA	NA					
v	$C_{38}H_{40}O_{20}$	$[M-H]^-$	815.20488	1.1	6.81	2298359	3.2E-06	0.00	X		Mixture of Isoorientin-sinapoylhexoside and isoscoparin-hydroxyferuloylhexoside	3	Piasecka
		$[M+H]^+$	817.21984	1.6	6.82	3784040	4.6E-05	0.00					
		$[M+Na]^+$	839.19982	-0.8	6.82	268433	NA	NA					
w	$C_{36}H_{36}O_{18}$	$[M-H]^-$	755.18293	0.0	6.83	1246	3.1E-04	0.49		X	*C*-hexosyl-luteolin-*p*-coumaroylhexoside	2	Dong, Gong
		$[M+H]^+$	757.19585	-2.1	6.83	20737	1.4E-03	0.63					
		$[M+Na]^+$	779.17734	-2.6	6.83	3713	2.6E-01	0.87					
x	$C_{27}H_{30}O_{15}$	$[M-H]^-$	593.15165	0.8	6.99	6174	2.2E-02	0.54		X	Isovitexin-2"-*O*-glucoside	2	Piasecka
		$[M+H]^+$	595.16466	-1.8	6.99	25418	1.8E-01	0.58					
		$[M+Na]^+$	617.14607	-2.6	7.00	2037	9.5E-01	0.88					
y	$C_{36}H_{36}O_{18}$	$[M-H]^-$	755.18403	1.5	7.20	1252	1.0E-02	0.44		X	*C*-hexosyl-apigenin-*O*-caffeoylhexoside	2	Dong, Gong
		$[M+H]^+$	757.19659	-1.1	7.19	43912	2.6E-03	0.51					
		$[M+Na]^+$	779.17876	-0.8	7.18	7012	1.0E-03	0.50					
z	$C_{21}H_{20}O_{11}$	$[M-H]^-$	447.09302	-0.6	7.19	6489	7.0E-03	0.37		X	Luteolin-7-*O*-glucoside, luteolin-5-*O*-glucoside or kaempferol 3-*O*-glucoside	2	Piasecka, Dong, Gong
		$[M+H]^+$	449.10709	-1.7	7.21	12458	3.1E-02	0.44					

Table 12, *continued*

Component	Formula	Ion	Peak *m/z*	Deviation *m/z* [ppm]	Peak RT [min]	Intensity (BH low) Ctrl	*p*-value	FC	B	R	Putative annotation	Annotation level	Reference
aa	$C_{39}H_{42}O_{20}$	$[M-H]^-$	829.22021	0.7	7.33	422932	7.5E-05	0.01	X		Isoscoparin 7-*O*-[6″-sinapoyl]-glucoside	2	Piasecka
		[M+Na-2H]-	851.20623	4.8	7.33	35859	1.8E-05	0.37					
		$[M+H]^+$	831.23290	-1.6	7.35	1294868	6.1E-04	0.01					
		$[M+Na]^+$	853.21467	-1.7	7.34	117247	NA	NA					
ab	$C_{38}H_{40}O_{20}$	$[M-H]^-$	815.20337	-0.8	7.48	37438	2.2E-05	0.23	X		Mixture of Isoorientin-sinapoylhexoside and isoscoparin-hydroxyferuloylhexoside	2	Piasecka
		$[M+H]^+$	817.22025	2.1	7.46	36630	NA	NA					
ac	$C_{37}H_{38}O_{18}$	$[M-H]^-$	769.19877	0.3	7.60	NA	NA	NA			Isovitexin-7-*O*-[X”-feruloyl]-glucoside or isovitexin-2”-*O*-[6”-(*E*)-feruloyl]-glucoside	2	Piasecka, Yang
		$[M+H]^+$	771.21198	-1.4	7.58	13352328	1.5E-03	0.67					
		$[M+Na]^+$	793.19322	-2.3	7.55	606174	3.1E-11	0.33					
ad	$C_{38}H_{40}O_{19}$	$[M-H]^-$	799.20901	-0.1	7.59	432674	2.2E-02	0.50	X		Isoscoparin 2"-*O*-(6'''-(*E*)-feruloyl)-glucopyranoside	2	Yang
		$[M+H]^+$	801.22109	-3.2	7.60	1323438	9.9E-04	0.34					
ae	$C_{37}H_{38}O_{18}$	$[M-H]^-$	769.19932	1.0	8.23	244839	5.4E-03	0.67	X		Isovitexin-7-*O*-[X”-feruloyl]-glucoside or isovitexin-2”-*O*-[6”-(E)-feruloyl]-glucoside	2	Piasecka, Dong
		$[M+H]^+$	771.21369	0.8	8.23	477823	8.3E-04	0.58					
		$[M+Na]^+$	793.19425	-1.0	8.23	115027	1.3E-09	0.23					
af	$C_{22}H_{22}O_{11}$	$[M-H]^-$	461.10918	0.5	8.12	63638	2.6E-02	0.45	X	X	Chrysoeriol-7-*O*-glucoside or chrysoeriol-5-*O*-glucoside	2	Piasecka, Dong, Gong
		$[M+H]^+$	463.12232	-2.5	8.14	122076	3.4E-02	0.50					
ag	$C_{23}H_{24}O_{12}$	$[M-H]^-$	491.11911	-0.8	8.16	187394	2.4E-02	0.78	X	X	Tricin-7-*O*-glucoside or tricin-5-*O*-glucoside	2	Piasecka, Dong, Yang, Matsuda
		$[M+H]^+$	493.13297	-2.2	8.18	411492	3.1E-02	0.75					
		$[M+Na]^+$	515.11395	-4.0	8.19	NA	6.6E-01	1.93					
ah	$C_{36}H_{36}O_{18}$	$[M-H]^-$	755.18424	1.8	8.25	NA	1.7E-02	0.64		X	Luteolin- or kaempferol-feruloylhexoside	3	–
		$[M+H]^+$	757.19621	-1.6	8.27	NA	4.5E-02	0.72					
		$[M+Na]^+$	779.17615	-4.1	8.26	25124	NA	NA					

4.5 Targeted analysis of phytohormones

Findings from transcriptomics analysis, performed by collaborators on *Lr34* barley, revealed the involvement of several genes related to the biosynthesis of the plant hormone jasmonic acid [27]. The involvement of hormones in plant defense is well known. Therefore, the plant hormones salicylic acid (SA, **25**), jasmonic acid (JA, **26**) and abscisic acid (ABA **27**) were investigated in barley and rice by a targeted analysis, for which a hormone specific extraction protocol was applied [186].



25 **26** **27**

Scheme 12: Plant hormones SA (**25**), JA (**26**) and ABA (**27**).

The targeted analysis of the plant hormones SA and ABA revealed trends of upregulation in high *Lr34*-expressing barley plants, whereas differences in rice were diffuse. For ABA and JA, hormone signals were below the detection limit for several samples, possibly due to low amounts of available plant material.

A strong accumulation of salicylic acid for all high-expressing genotypes of barley grown under all conditions including hydroponics, mock-infected and pathogen-infected conditions was observed (Figure 32). Also an oxidation product of salicylic acid, gentisic acid-*O*-glycoside, had been observed to be upregulated in high-expressing barley (component **o**, 4.4). For low-expressing *Lr34* barley plants, the SA (**25**) levels were similar to the control plants. In rice, a tendency of SA (**25**) upregulation was observed in high *Lr34*-expressing plants grown under mock- and pathogen-infected conditions. The concentration of ABA (**27**) was below the detection limit in some samples of mock- and pathogen-infected barley and rice plants (Figure 33). Nevertheless, clearly measurable ABA (**27**) concentrations were detected in high *Lr34* expressing barley plants grown under mock-and pathogen-infected conditions. High-expressing hydroponically grown rice even showed a decrease of ABA (**27**) concentration. At low expression level of *Lr34*, no clear differences of ABA concentrations were detected neither in barley nor in rice.

Investigation of JA (**26**) resulted in low signal intensities and the signals were below the detection level for many samples (Supplementary figure 29). Due to many missing values, no concise comparison between control and *Lr34* groups was possible.
In summary, these results suggest that the targeted hormone determination was not very sensitive. Due to limitations in the amount of plant material, extraction was performed with 50 respectively 30 mg of plant powder, which turned out to be insufficient for low levels of hormones. Another limitation was the quality of the plant material which was stored as plant powder at –70°C for an extended time period that probably lead to hormone decomposition.

In an alternative investigation, our collaborators observed a tremendous increase in SA concentrations of transgenic *Lr34* barley seedlings. Further, they also detected clear raise in levels of both jasmonic acid and jasmonic acid isoleucine [27]. From a biological point of view, the increase in jasmonic acid in context of *Lr34* makes sense, because jasmonic acid is well known to induce expression of a wide range of defense-related genes [169].
SA severely affects secondary metabolism, especially pathways that lead to phenolic compounds [187]. SA and ABA were reported to act antagonistically against each other [187]. In a study that applied single hormones to *Arabidopsis thaliana* cell cultures, increased concentrations of sugars and amino acids were observed upon ABA treatment and elevated malate and succinate levels were obtained after SA administration [187]. The combination of the effects observed for SA and ABA would fit to the results obtained for high *Lr34*-expression levels of barley, for which upregulations of both SA and ABA were observed, together with enhanced concentrations of the primary metabolites glucose, fructose, tyramine, leucine and succinate.

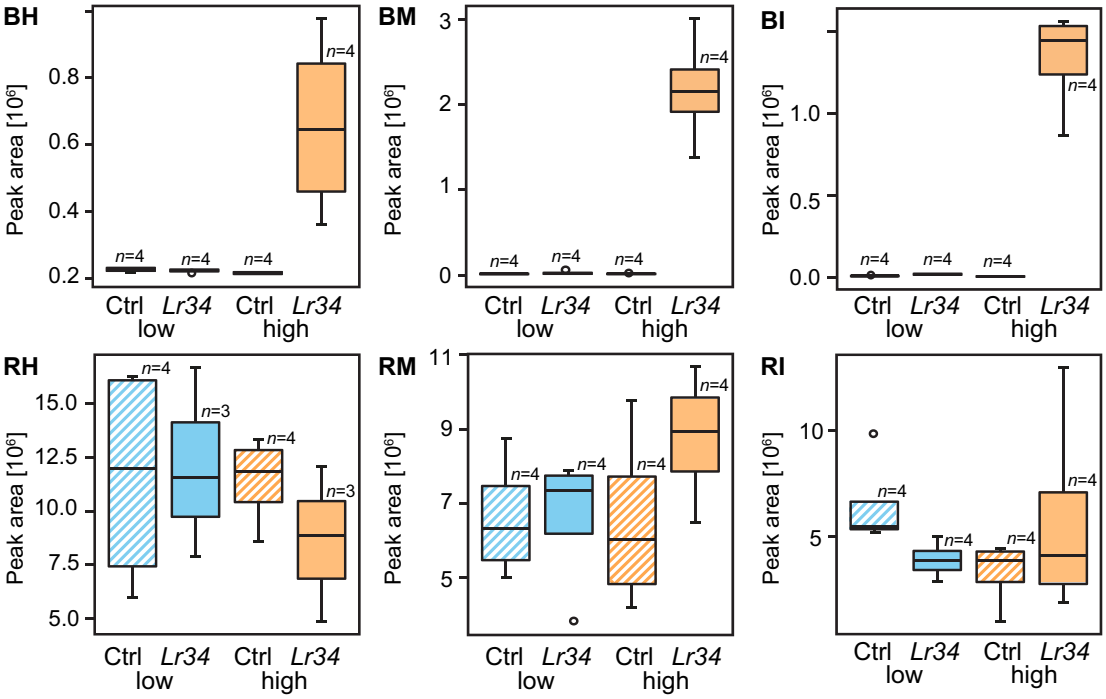


Figure 32: Signals of SA (**25**) in different experiments of barley and rice.

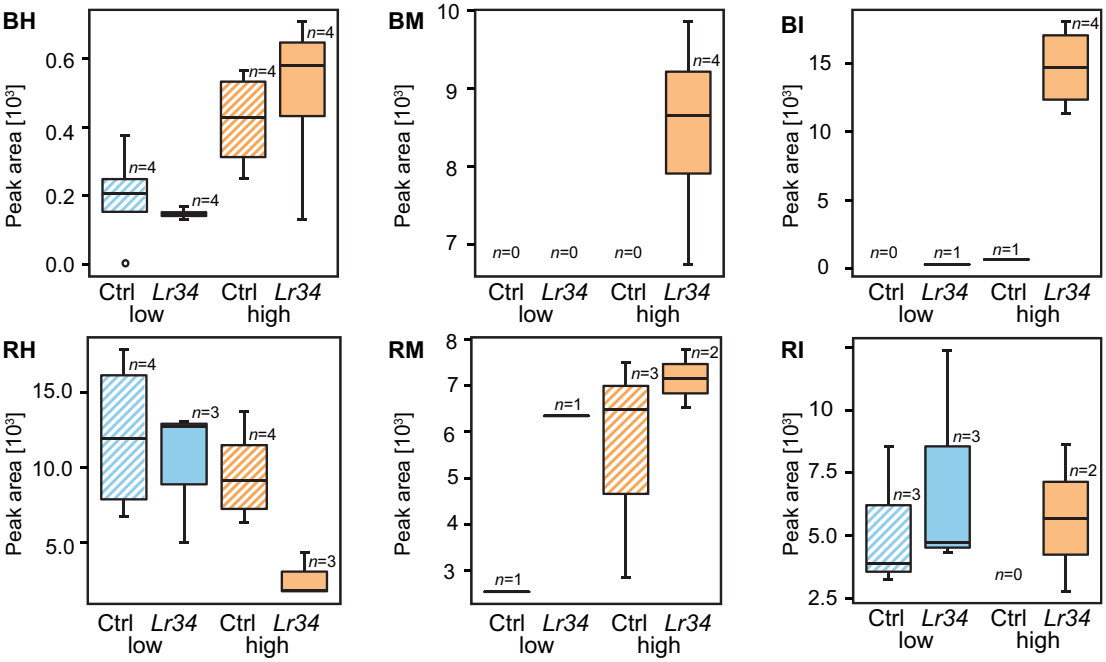


Figure 33: Signals of ABA (**27**) in different experiments of barley and rice.

4.6 Evaluation of hydroponic versus soil growth conditions

The initial hypothesis about the different growth conditions was that standardized conditions involving well-defined hydroponic growth medium would lead to less biological variation. Accordingly, the expectation was to find a higher number of significantly changed metabolic features in plants grown under hydroponic conditions than in the soil-grown plants of mock- and pathogen infection.

Table 13: Number of common metabolic differences for different growth conditions in barley and rice based on separate expression levels. Consistently regulated metabolites in plants grown under hydroponic conditions are highlighted in grey.

Low *Lr34* expression level	Barley			Rice		
	BH	BM	BI	RH	RM	RI
Primary metabolites	1	3	0	0	1	0
Lipids	0	0	0	0	0	0
LC-(–)-ESI-MS features of secondary metabolites	243	176	36	67	98	182
LC-(+)-ESI-MS features of secondary metabolites	20	118	3	2	34	39

High *Lr34* expression level	Barley			Rice		
	BH	BM	BI	RH	RM	RI
Primary metabolites	11	50	53	7	0	1
Lipids	33	69	43	0	0	0
LC-(–)-ESI-MS features of secondary metabolites	392	4018	5248	1154	702	331
LC-(+)-ESI-MS features of secondary metabolites	182	3347	5008	451	395	112

The numbers of common metabolic differences related to the *Lr34* gene that were detected within the different growth conditions in barley and rice are summarized in Table 13. The metabolic differences found for hydroponically grown plants are highlighted in grey, and they are compared to the growth conditions involving mock and pathogen infection in the following.

For rice plants with high *Lr34* expression level, more metabolic differences were detected in hydroponics than in mock- or pathogen-infected plants. In contrast, rice plants with low *Lr34* expression level showed fewer regulated metabolites when they were grown hydroponically than when they were grown on soil and mock-infected or pathogen infected. Thus, the initial hypothesis could be confirmed for rice for the high expression level, but not for the low expression level.

In high *Lr34* expressing barley, less features were detected to be significantly changed for the hydroponic growth condition than for the mock- and pathogen-infected conditions. This is in contradiction with our hypothesis.

For the low expression level in barley, the tendencies were not consistent for all classes of metabolites. One possible reason for the different trends observed in barley and rice was that the hydroponically grown barley plants were younger than the plants used for the mock- and pathogen-infected conditions (BH 14 d versus BM and BI 23 d or 28 d), whereas rice plants were almost the same age (RH 20 d versus RM + RI 23 d, Table 17, 6.1.2).

Together, these results suggested that barley and rice reacted differently to the growth conditions. Hydroponic growth medium was beneficial for the investigation of robust changes in primary and secondary metabolites in high expressing rice. In contrast, soil as growth medium was favorable for high expressing barley and low-expressing rice. Thus, no clear decision can be made with respect to which general growth condition should be used for future experiments.

4.7 Investigation of natural *Lr34* wheat under field conditions

In order to evaluate the metabolic profile of *Lr34* in its natural environment, we aimed at investigating field-grown plants. Transgenic *Lr34* barley and rice could not be grown under field conditions due to laws and regulations for the application of gene modified organisms. Therefore, the goal was to examine whether metabolic differences detected in barley and rice were also found in wheat. In wheat, two predominant *Lr34* alleles were found on chromosome 7D, of which one confers disease resistance while the other is susceptible [22]. The metabolomes of control wheat plants containing the susceptible allele were compared with the metabolomes of resistant *Lr34* wheat plants containing the resistant allele. For each group of control and *Lr34* plants, nine replicates were collected from individual plants.

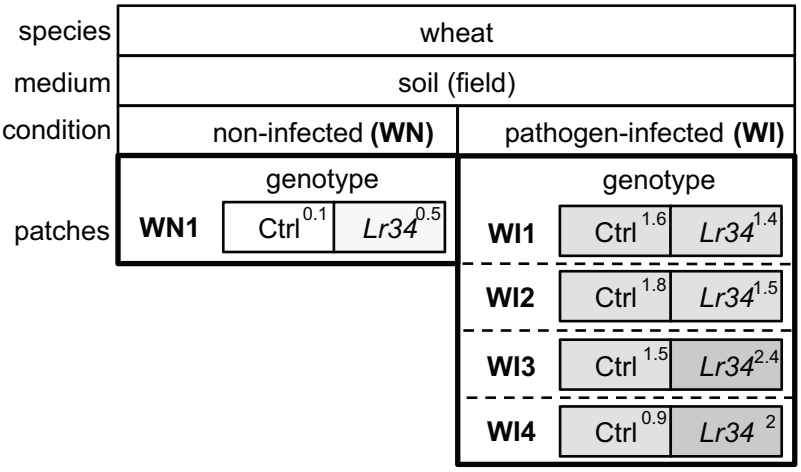


Figure 34: Experimental design for wheat grown under field conditions. *Lr34* plants were compared to respective control (Ctrl) plants. Patches are small experimental sub-fields within the field. Gray scale and numbers represent the estimated senescence in the harvested leaf (senescent part of the leaf relative to leaf length scaled to 10).

In wheat, *Lr34* is expressed at highest levels in the flag leaf [23], the last leaf developed in a mature plant. Therefore, whole flag leaves were harvested from field-grown wheat plants of the winter wheat cultivar *Arina* (Figure 34). Instead of replicate experiments, several small sub-fields within the field, so-called patches, were used for the pathogen-infected condition. Wheat plants in the field had been infected with wheat leaf rust (*Puccinia triticina* spores) 56 days before harvest by placing pots of strongly infected plants as sources of pathogen spores in the field. Additionally, non-infected samples of the same field were used for comparison. Details about plant material are documented in 6.2.1. The estimated senescence of the harvested flag leaves is shown as grey scale (Figure 34, Supplementary figure 30).

Basically, significant differences between control and *Lr34* wheat plants were investigated. In order to find common metabolic changes in non-infected and pathogen-infected wheat, a similar approach was used as previously described for barley and rice (Figure 35). First, metabolites that were significantly increased or decreased in their abundance in the four individual pathogen-infected patches WI1–WI4 were compared. Metabolites that were significantly regulated in all four patches WI1–WI4 (intersection marked with black cycle) and metabolites that were intersecting amongst three patches (marked with grey circle) were selected for the comparison with differences observed in the non-infected patch WN1.

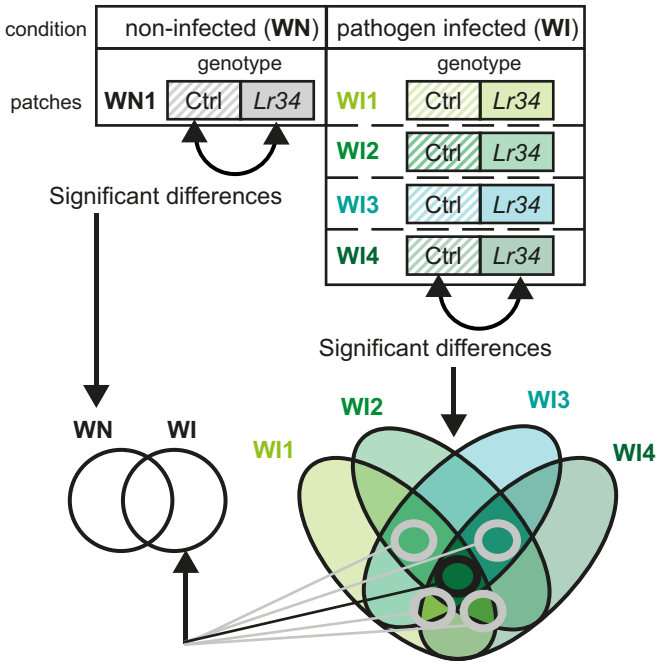


Figure 35: Common metabolic differences in wheat: common significantly different metabolites in pathogen-infected patches WI1–WI4 were compared to differences from the non-infected patch WN1. Intersections of 4 groups (black circle) as well as 3 pathogen-infected patches (grey circles) were taken into consideration.

In analogy to barley and rice, the primary metabolites and lipids were identified prior to statistical analysis. In wheat, 86 primary metabolites (Supplementary table 3) and 156 lipids (Supplementary table 5) were identified. In contrast, features of secondary metabolites were statistically analyzed and compared without knowing their identity.

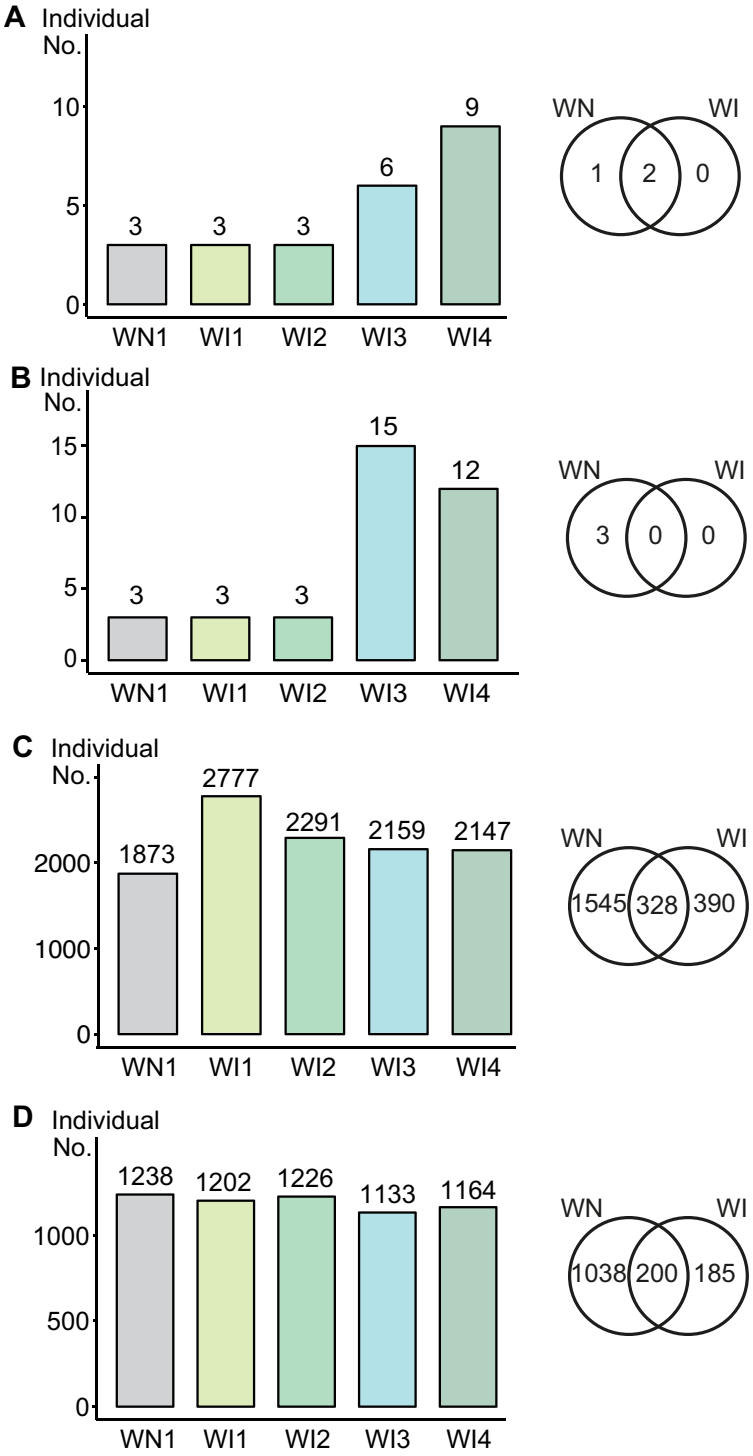


Figure 36: Significantly different metabolites in individual wheat experiments and overlap of non-infected and pathogen infected conditions: primary metabolites (A), lipids (B), secondary metabolite features detected in LC-(–)-ESI-MS (C) and LC-(+)-ESI-MS (D).

Comparisons of the significant metabolites detected in non-infected and infected wheat are shown for the different metabolite classes in Figure 36. In agreement with the data from barley and rice, the highest number of differences was obtained for the secondary metabolite fingerprints.

For all comparisons of control with *Lr34* wheat, pathogen-infected or not, the same two primary metabolites were found to be different: while succinate (**17**) was upregulated, isocitrate (**15**) was downregulated (Figure 37). Previously, the same trend had been observed for rice. The decrease in isocitrate (**15**) concentration was associated with the glyoxylate cycle (Scheme 6).

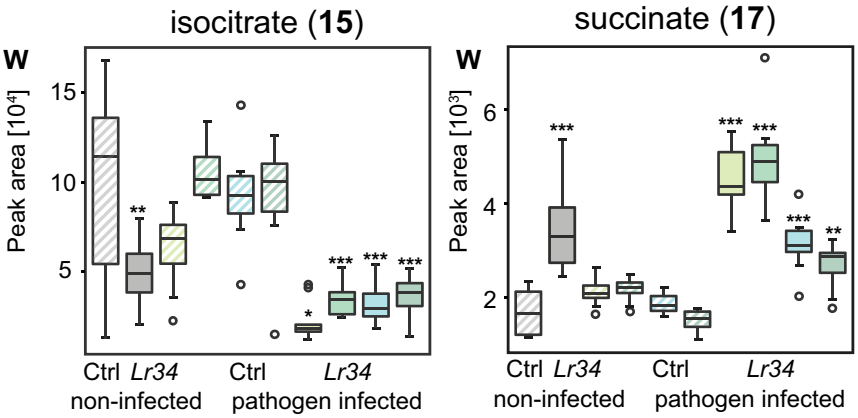


Figure 37: Primary metabolites involved in glyoxylate cycle in wheat.

In terms of lipids, no joint difference in lipid accumulation could be detected for wheat grown without and with pathogen infection. For non-infected wheat, three upregulated lipids (Cer t18:1/c24:1, MGDG 36:5(1), and DGDG 36:2) as well as a single downregulated lipid (MGDG 38:6) were detected.

The investigation of secondary metabolite features produced high intersections of simultaneously regulated features in conditions without and with pathogens. For both conditions, 328 and 200 features were detected to be commonly regulated in the LC-(–)- and LC-(+)-ESI-MS, respectively. So far, the intensity of the responsible secondary metabolites had not been investigated.

The targeted analysis of the plant hormones salicylic and abscisic acid showed no clear trend for *Lr34* wheat (Figure 38), and jasmonic acid was only detected in few samples (Supplementary figure 29).

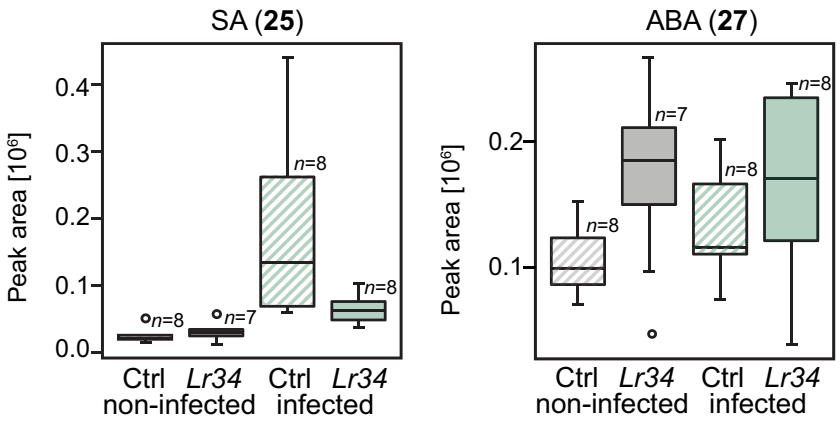


Figure 38: Signals of SA and ABA in non-infected and pathogen-infected wheat.

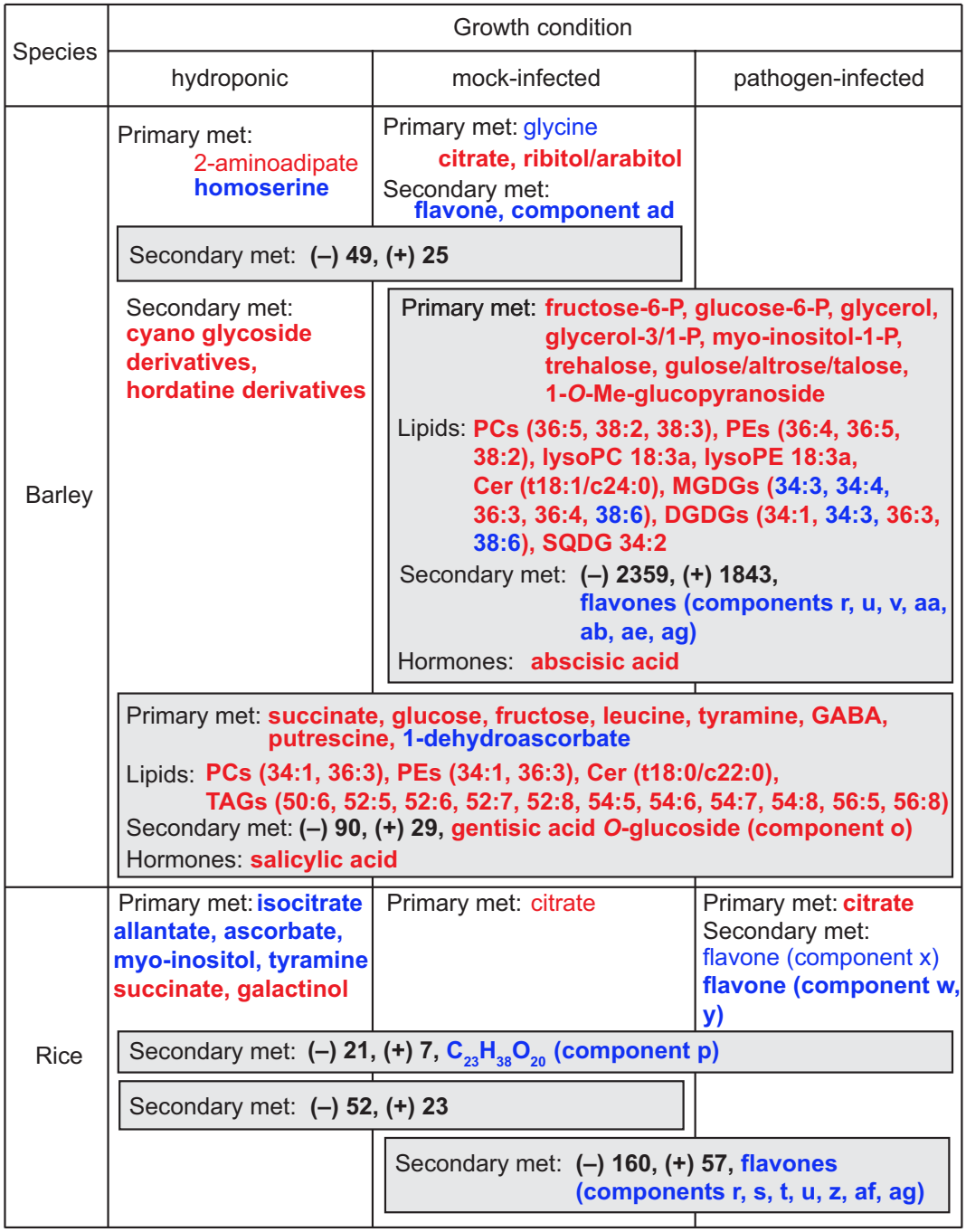
In summary, the investigation of field-grown wheat confirmed some of the metabolic trends observed rice. Concerning primary metabolites, the induction of the glyoxylate cycle detected on metabolite and transcript level in rice was confirmed in wheat in terms of isocitrate downregulation and succinate upregulation. Similar to the observation in rice, no robust regulation of lipids was observed in both non-infected and pathogen-infected wheat. Also the targeted analysis of the plant hormones SA and ABA showed no clear trend in *Lr34* wheat, which was as well in accordance rice. The highest change was obtained for secondary metabolites, which were not annotated so far and thus remain unknown. The only similar effect that was observed in wheat and in barley was the lipid MGDG 38:6, which was downregulated in non-infected wheat and in high expression levels of mock- and pathogen infected barley. Thus, the metabolic response of *Lr34* wheat observed so far mainly resembled the one in rice.

4.8 Discussion

Metabolomics captures a snapshot of the molecular phenotype of a biological system. In this study, we aimed at profiling the metabolic responses of the durable, multi-pathogen resistance *Lr34*. In this extensive investigation, three different cereal species (barley, rice and wheat) were investigated under different growth conditions (hydroponic, mock / or non-infected and pathogen-infected growth). For barley and rice, metabolic profiles of plants with two expression levels of *Lr34* were analyzed. To overcome limitations in metabolic coverage, a comprehensive protocol was chosen with fractionated extraction and combined GC- and LC-MS technology allowing the simultaneous analysis of primary and secondary metabolites as well as lipids from a single sample. Targeted plant hormones were analyzed by a separate extraction. Although not entirely comprehensive, this applied metabolomics approach allowed assessing the basic classes of metabolites.

For primary metabolites and lipids, metabolite identification was performed prior to the statistical analysis, and the statistical relevance was only evaluated for the identified metabolites. In contrast, all LC-ESI-MS features of the secondary metabolite fingerprints were used for statistical analysis, and the metabolite identification was focused on interesting components that were changed between *Lr34* and control plants grown under different conditions.

The main metabolite changes associated to *Lr34* for barley and rice are summarized in Figure 39. Generally, the majority of metabolic differences occurred in high expressing genotypes of barley, including primary metabolites, lipids, secondary metabolites and hormones. In rice, no common metabolic changes between the three growth conditions were observed for primary metabolites or lipids, but in the secondary metabolic fingerprints of high *Lr34* expressing genotypes. In wheat, the primary metabolites isocitrate and succinate were consistently regulated, besides a few lipids and a high number of secondary metabolic features.



Species	Growth condition	non-infected	pathogen-infected
Wheat		Primary met: succinate, isocitrate Lipids: no common changes detected Secondary met: (–) 328, (+) 200 Lipids: Cer (t18:1/c24:1), MGDGs (36:5(1), 38:6), DGDG 36:2	

Figure 39: Summary of metabolic changes observed in *Lr34* in this work: upregulated metabolites are shown in red, downregulated metabolites in blue; metabolic changes in plants with high and low *Lr34* expression levels are indicated in bold and regular font, respectively. Abbreviations: met: metabolite, (–): LC-(–)-ESI-MS, (+): LC-(+)-ESI-MS

Metabolic differences between control and *Lr34*

The metabolomic datasets had been visualized for each metabolite class by principal components. Basically, the same tendency was observed for primary metabolites, lipids and secondary metabolites: primarily, plants with high *Lr34* expression level were separated from the respective controls and from low expression level, and the effect was most prominently observed for barley. This indicated a strong influence of the senescence phenotype on the metabolome.

The design of the investigation for the metabolomic characterization of *Lr34* would also allow studying the principal metabolic response of the different cereal species barley, rice and wheat on pathogen infection, based on the control plants. Considering the amount and complexity of the current metabolomics approach, we focused at finding common differences in the metabolic profiles between control and *Lr34* resistant plants grown under different conditions.

Our strategy objected at finding robust metabolic changes associated with *Lr34* that occurred in plants grown under the three different conditions involving hydroponics, mock- and pathogen infection and that were conserved between two experiments performed for the same growth conditions. Because the leaf-tip necrosis phenotype was observed in all growth conditions, especially in barley, the hypothesis was that there were common metabolic responses in all growth conditions that finally lead to *Lr34* resistance.
To evaluate metabolites or metabolic features that were regulated in plants grown under different conditions, two strategies were applied, considering either (A) combined expression levels and replicate batches, or (B) the low and high expression levels separately.
For barley and rice, only few primary metabolites and lipids were found to accumulate at different levels in transgenic *Lr34* plants grown under different conditions. Not surprisingly, more significant changes were detected within secondary metabolite features. While primary metabolites and lipids were restricted to the identified number of metabolites (84 primary metabolites and 156 lipids), all detected LC-ESI-MS of features were considered for the evaluation of changes in secondary metabolism. Furthermore, the general fold changes in secondary metabolism tend to be higher than in central metabolism [188]. Additionally, many features might represent various adducts and fragments of a single metabolite. In plants with low *Lr34* expression levels, only few secondary metabolites were

consistently regulated in plants grown under different conditions. This result might indicate that the individual growth conditions promote dynamically changing secondary metabolites.

Considering the common metabolic differences described in this study, various biological, methodological and strategic factors influenced the final number of detected differences. The biological factors involved the growth conditions and the natural variability of metabolites in individual replicate plants, reflected by the significance of metabolic differences (group variability, fold change). Methodological factors influencing the detected metabolic differences were e.g. the range of detected metabolites (coverage) or the metabolite concentration (detection limit, stability). Strategic factors were for example the decision to evaluate only metabolic changes that were found in both individual batches of the same growth condition (robustness), or the decision to prioritize intersections of regulated metabolites in plants grown under different growth conditions.

Looking at the initial number of metabolic changes detected in the individual growth batches listed in Supplementary table 8, it was recognized that generally the number of significant changes was low, indicating a high variability between individual replicate plants or a low fold change. A comparison of the number of features found in both batches with the number of intersecting features revealed that only a few were consistently regulated in both batches, especially for plants with low *Lr34* expression level. Thus, the metabolic fingerprints varied between replicate batches. Possible reasons for the variations are discussed in the following.

Changes in the environmental conditions, for example caused by temperature fluctuations or light variations are influencing the plant metabolome [189]. Because the plants were grown in phytotron climate chambers or in environmentally controlled greenhouses, the abiotic factors were expected to contribute less to the variation between individual batches.

Microbes that are present in the growth medium — soil or hydroponic solution — may negatively influence a metabolomics experiment [190]. They represent a possible source of variation of the metabolic profiles and fingerprints between the individual batches.

For barley BM and BI, the two batches were harvested at different time periods after inoculation due to slower development of the leaf rust infection symptoms, which caused a shift in plant growth and development stage. The developmental stage had been reported to influence metabolite levels [191], and therefore part of the variation between individual batches might be attributed to the different plant developmental stage at harvest.

Furthermore, levels of certain metabolite levels change within the day/night cycle [192]. Because the harvesting time point within the day was different for the two batches of mock- and pathogen-infected barley (BM1/BI1 16 h versus BM2/BH2 11 h, Supplementary table 9), some of the inter-batch variation might also be attributed to the diurnal changes of plant metabolism.
The inoculation process for pathogen infection might contribute to plant stress, e.g. by the plastic hood used to protect the sprayed plant [190]. Therefore, plants grown under mock- or pathogen-infection might have developed stress symptoms that were influencing the inter-batch variation on the metabolite levels.

In summary, the metabolite levels were shown to vary between individual batches, and highlight the importance of the experimental setup for good reproducibility, including the controlled environmental conditions, the developmental stage of the plant and the harvesting time point during the day.
The current results obtained in barley and rice suggests to use even tighter controlled growth conditions to improve the information value gained by metabolomic characterization of *Lr34*. On the other hand, metabolic differences could also be observed for field grown *Lr34* wheat, where the environmental variance for plant growth was high. So far, secondary metabolite differences observed in *Lr34* wheat were not yet annotated, but this might be worthwhile. For the investigation of metabolic differences in primary metabolism, a higher number of samples might be necessary to observe significant effects.
To monitor the development of fungal infection symptoms, alternative quantitative methods are suggested for future experiments. Quantitative methods for the evaluation of the extent of fungal infection involve *in planta* staining of chitin combined with fluorescence measurement [193] or the detection of fungal biomarkers like steroid ergosterol derivatives [194] or arachidoic acid [43].

Initially, we hypothesized that standardized conditions involving well-defined hydroponic growth medium would lead to less biological variation. Comparison of the number of significantly changed metabolic features in plants grown under hydroponic conditions with the ones from the soil-grown plants of mock and pathogen infection (4.6) revealed that barley and rice reacted differently to the growth conditions. While hydroponics was favorable for robust changes in primary and secondary metabolism in high expressing rice, soil as growth medium was beneficial for high expressing barley and low-expressing rice.

Therefore, no clear suggestion for a general growth condition of subsequent experiments could be obtained.

Metabolite annotation of robust metabolic differences of secondary metabolites

Primary metabolites and lipids were identified prior to statistical analysis. The identification of primary metabolites and lipids relied on libraries of reference compounds or previously elucidated lipids. After careful manual evaluation, a variety of compound classes were annotated. Identified primary metabolites included amino acid derivatives, organic- and fatty acids, and sugars. Lipids belonging to classes of phospholipids (PG, PI, PC, lysoPC, PE, lysoPE, PS), sphingolipids (Cer, GlyCer), galactolipids (MGDG, SGDG, SQDG) and triacylglycerols were successfully identified.

The majority of the secondary metabolites in plants are unknown. Therefore, only robustly significant differences of secondary metabolic fingerprints were selected for metabolite annotation. Secondary metabolite identification in complex matrices still comprises a major bottleneck in metabolomics, and the difficulties in metabolite identification based on LC-MS arise from various challenges like the huge chemical diversity of plant metabolites, the variability of MS/MS spectra depending on instrumental conditions and the lack of commercially available standards for a wide range of plant metabolites.
An initial step towards metabolite identity is the calculation of the chemical formula. Multiple features were often significantly regulated for a single component of interest. The features comprised adducts and in-source fragments, but also fragments of unknown origin, and it was difficult to allocate the features to mass shifts of known adducts. To find out if a feature possibly corresponded to a fragment of the metabolite, the HR-MS spectra of the respective retention time was investigated manually and the peak shape of EICs of different features were compared. In order to limit the possible elemental compositions of a chemical formula, relative isotope abundances were considered. For low abundant features, relative isotope abundances were often interfering with mass peaks from co-eluting metabolites.
Another difficulty was the acquisition of informative MS/MS fragment spectra to gain further structural evidence. To select an appropriate precursor ion, the molecular mass should be known. Further, the collision energy needs to be adjusted to yield a variety of fragments of sufficient intensities. Generally, the isolation of precursor ions might be difficult for low abundant ions or reactive metabolites that possibly degrade during storage of the plant extracts.

For barley and rice, part of the secondary metabolites that were significantly regulated in transgenic *Lr34* plants were putatively annotated. Based on MS/MS fragmentation behavior and similarities to literature, gentisic acid glycoside and two structurally related tetrasaccharides were found in barley and rice, respectively. Further, a group of *C*-glycosylated flavones was putatively identified in barley and rice based on reference spectra found in literature, thereby allocated to metabolite identification level 2.
Although a number of secondary metabolites were putatively annotated, the metabolite identity of a high number of prominent secondary metabolites responsible for the differences between control and *Lr34* plants remained unknown.

Several strategies have been developed to overcome the issue of limited secondary metabolite identification. Plant growth on multi-isotopically labeled media is a powerful approach for precise chemical formula calculation and assists metabolite annotation, but suitable growth chambers are required for $^{13}CO_2$-labelling. Compiling an in-house library of commercially available reference compounds allows metabolite identification with high certainty but requires considerable effort and cost, because spectra usually need to be recorded at several collision energies with standardized protocols. While commercial services are profiling around thousand metabolites of the human metabolome, plant metabolite diversity is still sparsely covered with commercially available reference compounds. Another approach involves the purification of natural products for NMR structure elucidation; a recent study used 1.4 kg of barley leaves and was able to identify 152 metabolites [74]. The advantage of plants is that the production of plant material can be up-scaled in order to isolate specific metabolites, which is not possible for biomarkers of human disease. However, the procedure of isolation still requires extensive labor and is usually only applied to compounds of high interest. In this work, the major limitation was the interpretation of the data starting from selecting the right feature to calculate the sum formula, interference with coeluting ions for applying relative isotopic abundance algorithms, and also the recording and interpreting of MS/MS spectra.

Metabolic context of *Lr34*

The extent of leaf-tip necrosis induced by *Lr34* depends on the species, the *Lr34* expression level and on environmental conditions. In barley, a strong leaf-tip necrosis-type senescence was observed that ultimately affected entire leaves, together with negative effects on plant growth and fitness. The barley phenotype was much stronger than in rice,

where leaf-tip necrosis was restricted to leaf tips and resembled the *Lr34* phenotype of wheat.

Another example of a resistance resulting in a growth penalty is the *mlo* resistance gene in barley, which provides durable resistance against powdery mildew. The *mlo* gene encodes a transmembrane protein that interacts with calmodulin. Loss-of-function mutations of *Mlo* were a major success in barley breeding for nearly 40 years, although agronomic costs associated with *mlo* alleles in barley involve necrotic spotting, yield penalty and susceptibility to other facultative pathogens like rice blast or *Fusarium* head blight [195].

A further example of a multi-pathogen resistance gene in wheat is *Lr67,* also providing partial resistance to leaf-, stripe- and stem rust as well as to powdery mildew. Introduction of *Lr67* into wheat showed a similar phenotype than *Lr34* including the pleiotropic leaf-tip necrosis [196], and was also functionally transferable to barley [21]. In field trials, no effects of *Lr67* were observed on agronomic or quality traits. The corresponding LR67 protein was smaller (514 amino acids) with twelve transmembrane helices and similarity to a family of H^+/monosaccharide symporters facilitating hexose transport across the plasma membrane [21]. The resistant alleles were characterized by two single-nucleotide polymorphisms, of which one was found to be primarily responsible for the loss-of-function. Yeast transport assays confirmed high-affinity uptake of ^{14}C-glucose and a lower affinity to fructose by the susceptible LR67, while the resistant version was unable to import glucose. Protein plasma membrane localization was confirmed as well as homo- and heterodimer formation, and the resistance mechanism may be caused through forming inactive hetero-multimeric protein complexes. Resistant *Lr67* is thought to block hexose regeneration from the apoplast, consequently increasing the hexose/sucrose ratio in the leaf apoplast. As a result, sugar-mediated defense responses may be induced. Sugar perception in plants is highly complex; they not only contribute to physiological processes and signaling during plant defense but also appear to limit biotrophic fungal pathogen infections.

One of the important results of the current thesis is the finding that the highest number of metabolites with significantly altered concentrations in *Lr34* plants was found for barley plants with high *Lr34* expression levels. Barley plants with high expression levels also showed the strongest senescence phenotype. Therefore, the strong leaf-tip necrosis-type senescence in barley clearly affected the metabolome. In terms of lipids, the upregulated sphingolipid Cer t18:0/c22:0 might be associated with the regulation of cell death, and upregulated TAGs might be formed as intermediates of galactolipid degradation in

senescent leaves. The decrease observed in levels of various flavones could also possibly be related to senescence. The interpretation of the results obtained by metabolite profiling and fingerprinting was therefore complicated by the strong phenotype. Direct effects caused by the unknown substrate of the LR34 transporter could not be differentiated from indirect secondary effects of the extended leaf tip necrosis phenotype. One hypothetical possibility to study the early effects of *Lr34* on metabolite level would be to use inducible cell lines, which could be investigated in a time-resolved manner.

The metabolomic analysis of *Lr34* barley revealed increased levels of glucose and fructose. In rice and wheat, the metabolites isocitrate and succinate pointed towards an increased glyoxylate cycle, in congruence with transcriptomics data. The glyoxylate cycle allows the synthesis of carbohydrates from lipids via acetate obtained in fatty acid *β*-oxidation, but both metabolite classes — lipids and carbohydrates — were not observed to be increased in whole-leaf extracts of rice and wheat. But still, there might be an indirect secondary connection to sugar-mediated defense responses observed in the multi-pathogen resistance *Lr67* with similar phenotype.

Outlook

To further explore the comprehensive dataset obtained in this work, additional effort could be attributed to the annotation of secondary metabolites in wheat.

All metabolomics experiments related to *Lr34* described in this chapter were based on whole leaves. Because *Lr34* encodes an ABC transporter protein, the initial metabolic changes caused by modified *Lr34* transport activity in the resistant plants might not be observable on whole leaf level, where no information about metabolite concentrations in sub-structural compartments could be obtained. Currently, the location of LR34 remains unknown, but knowledge about the localization of the transporter would definitely enhance the understanding of the *Lr34* resistance mechanism. For membrane proteins, localization approaches by fluorescence labeling approaches are generally difficult to implement.
In the following, approaches to investigate the metabolome of *Lr34* plants in a compartmentalized manner are discussed.
Assuming that LR34 is located in the plasma membrane, an interesting option would be to investigate the apoplast — the extracellular continuum outside cell membranes — and to use apoplastic fluid for metabolic fingerprinting. This approach would likely reduce the complexity of the observed matrix. Although collection of apoplastic fluid is well

established by vacuum infiltration centrifugation technique for leaves of e.g. tomato [197], it is more difficult for rigid *Poaceae* leaves like wheat [198], barley [199] and rice [200]. As quality control, leakage of symplastic compounds during extraction should to be tested, e.g. by malate dehydrogenase assay [199]. Furthermore, it might be difficult to collect sufficient apoplastic fluid for the purpose of metabolomics. Generally for metabolomics experiments, rapid sampling techniques are preferred because active enzymes alter the metabolome during complex sample collection procedures such as the collection of apoplastic fluid.
Another hypothetically possible option would be to transform *Arabidopsis* as cell culture cells with *Lr34* and to use the exudate for metabolic fingerprinting. However, the dilution of exudates in the medium and interference by culture medium components might complicate this approach.

The location of LR34 in an intracellular membrane might be another possibility. Metabolic profiles of separate sub-cellular compartments can be investigated by non-aqueous fractionation, where organelles are fractionated in a continuous density gradient before metabolic profiling [201]. Disadvantages of this method are the larger amount of required plant material (4–8 g) and the more complex sample preparation.

Compared to the susceptible LR34 transport protein, the deletion of a single phenylalanine postulated in a transmembrane domain was found to be required and sufficient for *Lr34* resistance in barley [27]. Therefore, the resistant variant of the LR34 transporter is hypothesized to either differ in molecular specificity, binding affinity for the substrate or binding affinity of the transporter to an unknown interacting protein. Besides potentially being a metabolite, the substrate(s) could also be ions or small proteins. However, based on the current hypothesis, the more specific approaches of metabolite analysis described above might help to find a substrate candidate of the LR34 transporter.

A further option to learn more about *Lr34*-mediated resistance involves the combination of the obtained detailed data sets from metabolic profiling with other findings. Further data e.g. from detailed physiological characterization of the phenotype and genetic approaches of *Lr34*-containing species grown under different conditions could be used to identify the underlying pathway(s) responsible for the phenotype. Together, this would allow to gain a more comprehensive understanding of the *Lr34* mediated resistance mechanism.

In a next step, recombinant expression of LR34 in yeast strains and plant protoplasts or plant cell cultures would be desirable for functional analysis to test transport capacity, kinetics and competition of candidates. Investigation of the molecular nature of the transported molecule(s) is clearly an essential step towards the understanding of *Lr34* durable resistance. Additional methods to study the mechanism of *Lr34* possibly involve co-crystallization of the transporter with the substrate, protein-ligand interactions studies or modeling.

With additional knowledge gained by these studies, the molecular and metabolic consequences of the substrate that contribute to multi-pathogen resistance could be interpreted with more clarity.

5 Conclusion

The *Lr34*-based resistance is of high value to plant breeding because of its broad effectiveness and durability, but also as a model to expand the knowledge on durable pathogen resistance. Interpretation of the molecular consequences caused by modified *Lr34* transport characteristics are intrinsically difficult, as the metabolic interplay is highly complex and dynamic. In this work, we observed distinct responses in different species; while primary metabolite response in rice and wheat was similar (e.g. reduced isocitrate levels), barley exploited a general upregulation of many metabolites, presumably also related to the strongly pronounced phenotype (Figure 39).

This work for the first time presented a broad characterization of the metabolic response of *Lr34* in terms of primary-, secondary metabolites, lipids and plant hormones in different cereal species. The metabolomics studies presented here form the basis for future work addressing the roles of various metabolites in the metabolic response of *Lr34* plants under different growth conditions.

The molecular analysis of *Lr34* resistance may provide insight into pathogen virulence strategies and host basal defense mechanisms. Finally, the precise chemical nature of the transported molecule(s) by *Lr34* as well as the substrate's detailed role in mediating resistance to the multiple pathogens in the different species needs to be further characterized.

6 Experimental procedures metabolomic characterization of *Lr34*

6.1 Metabolic fingerprinting of transgenic *Lr34* barley by LC-MS

6.1.1 Chemicals and reagents

Mass spectrometry grade acetonitrile (MeCN), methanol (MeOH) and formic acid (99.995%) as well as ampicillin (CAS 69-53-4) and camphor sulfonic acid (CAS 46365-22-4) in analytical standard quality were purchased from Sigma–Aldrich (Buchs, Switzerland). Water was obtained from a Millipore high purity water dispenser (Billerica, MA).
For hydrolysis experiments, apigenin (CAS 520-36-5), genkwanin (CAS 437-64-9), naringenin (CAS 67604-48-2), sakuranetin (CAS 2957-21-3), vitexin (CAS 3681-93-4) and hesperetin (520-33-2), ascorbic acid (CAS 50-81-7), were purchased from Sigma–Aldrich (Buchs, Switzerland), hydrochloric acid 33% (CAS 7647-01-0) from Thommen–Furler (Rüti b. Büren, Switzerland).

6.1.2 Plant material

Stable transgenic lines of barley (*Hordeum vulgare*) cv. Golden Promise were selected for metabolic profiling. Genomic sequence of *Lr34* was introduced to barley by *Agrobacterium tumefaciens*-mediated transformation. Specifically, two independent transgenic events expressing full-length *Lr34res* sequence were used, namely the low-expressing *Lr34res* line BG8 and the high-expressing *Lr34res* line BG9 [27]. *Lr34res* transcript levels in BG9 were found to be about 2.5 times higher than in BG8. While BG9 was found to have a single insertion, BG8 was found to have two copies [25]. Sister line plants that are identical to the transgenic lines apart from the *Lr34* genetic background were used as control.

6.1.3 Hydroponic plant growth

Barley seeds were surface sterilized with sterilization solution (sodium hypochlorite 1%, Triton-X-100 0.03%) for 20 min and stratified for 3 d at 4 °C. Seeds were pre-germinated vertically in square petri plates (Greiner, Frickenhausen, Germany) for 5 d in a climate chamber. Young seedlings were transplanted into a hydroponic system (General Hydroponics, Sebastopol, USA) filled with nutrient solution based on Hoagland nutrient solution: K_2SO_4 2.5 mM, $Ca(NO_3)_2$ 5 mM, KH_2PO_4 0.5 mM, $MgSO_4$ 2.5 mM, KCl 0.25 mM, H_3BO_3 2.5 µM, $MnSO_4$ 2.5 µM, $CuSO_4$ 0.5 µM, $(NH_4)Mo_7O_{24}$ 0.05 µM, $ZnCl_2$ 2.5 µM, Fe-Na-EDTA 0.25 mM, pH 6.5 by KOH 1 M. Plants were grown at light cycles of 16/8 h

light/dark in phytotron climate chambers with 450 µM mercury discharge light at 70% humidity, 20 °C day / 16 °C night temperature. Whole leaves were harvested and flash frozen in liquid nitrogen. After around 16 days, plants were in growth stage 1.4 (fourth leaf at least 50% emerged), starting first tillers.

6.1.4 Plant growth for hydrolysis

Barley sister line (Sib) and transgenic *Lr34* line BG9 were grown on soil in a environmentally controlled green house. Three replicates were harvested for both Sib and BG9, each whole mature sixth leaves of two different plants, and lyophilized. While sister line leaves were fully green, BG9 leaves were partly senescent and only around half the size. Leaves were weighted and ground in cryomill in steel vessel (30 Hz, 5 min at ambient temperature). Of the plant powder, 50 mg was weighed in 1.5 ml Eppendorf tube and kept in dessicator until hydrolysis.

6.1.5 Extraction

Within one minute, leaves were ground manually to fine powder with mortar and pestle that were chilled with liquid nitrogen (porcelain, 6.5 cm and 11.5 x 2.4 cm respectively), and 100 ± 2 mg of plant powder was weighted into 1.5 ml micro screw tubes (polypropylene with O-ring, Sarstedt, Nümbrecht, Germany). At the day of LC-MS analysis, ice-chilled solvent (MeOH 99.875%, formic acid 0.125%, ampicillin 0.25 mg l^{-1}, camphor sulfonic acid 0.25 mg l^{-1}) was added to the plant powder in a ratio of 1:3 [w/v] (300 µl), vortexed and sonicated 15 min at 20 ± 3 °C. After centrifugation (5 min, 20817 g, 4 °C), pools of each control / transgenic group and a quality control (QC) pool of all samples were prepared and vortexed. Samples and pools were diluted 1:1 [v/v] with ultrapure water, centrifuged, and 100 µl aliquots were transferred to LC-MS vials and stored in the auto sampler.

6.1.6 LC-MS analysis of plant extracts

An *Acquity UPLC* (Waters, Baden-Daettwil, Switzerland) was used for the inlet, controlled by *HyStar* chromatography software (Bruker Daltonics, Bremen, Germany). Barley leaf extracts were separated using a (MeCN)/ultrapure water solvent system containing 0.1% [v/v] formic acid.

The short method applied a flow rate of 0.2 ml min^{-1} on an *Acquity* BEH C18 reversed phase analytical column (1.7 µm, 1 x 50 mm, guard column 2.1 x 5 mm, Waters) maintained at 30 °C, with a binary gradient from 0–0.5 min at 3% MeCN, 0.5–8.5 min

increase 3–100% MeCN, 2.5 min at 100% MeCN and re-equilibration for 3 min at 3% MeCN.

For the long chromatographic method a different column geometry was used: *Acquity BEH* C18 column (1.7 µm, 2.1 x 100 mm, guard column 2.1 x 5 mm, Waters) maintained at 40 °C. Sample volumes of 5 µl were injected with partial loop injection mode, with needle washing with 600µl of weak (MeOH : water 1:9 [v/v]) and 400 µl of strong wash (MeOH : water 1:1 [v/v]). The following linear binary gradient was applied at a flow rate of 0.4 ml min^{-1}: 0–0.5 min at 3% MeCN, 0.5–13 min from 3–25% MeCN, 13–23 min from 25–100% MeCN, wash for 7 min at 100% MeCN with 0.6 ml min^{-1}, re-equilibration 5 min at 3% MeCN. Injections were performed in randomized order, with 2 QC pools in the beginning and after every 7 injections.

The LC was coupled to a *maXis* ESI-QTOF mass spectrometer (*Bruker Daltonics*) operated both in positive and negative electrospray mode. Instrument settings for positive ionization were as follows: 4000 V capillary voltage, –500 V endplate offset, with a N_2 nebulizer pressure of 2.0 bar and dry gas flow of 10 l min^{-1} at 200 °C, transfer funnel RF of 200 Vpp, multipol RF of 100 Vpp, quadrupole ion energy of 3.0 eV, collision cell collision energy of 5.0 eV, transfer time of 100 µs, collision RF of 100 Vpp, and 5 µs prepulse storage. In negative ESI, N_2 nebulizer pressure of 4.0 bar was used, with multipol RF of 200 Vpp and collision cell collision energy of 10.0 eV. MS acquisitions were performed in the full scan mode in the mass range from *m/z* 80–1200 at 25000 resolution (full width at half maximum) and 2 scans per second. Mass calibration was performed in dead volume of each chromatographic run, using sodium formiate clusters (10 mM solution of sodium hydroxide in isopropanol : water 1 : 1 [v/v] containing 0.2% formic acid) over a mass range of *m/z* 90–1178 (17–point calibration). As lock masses, methyl stearate ($C_{19}H_{39}O_2$) and hexakis ($C_{12}H_{18}N_3O_6P_3F_{12}$, $C_{18}H_{18}N_3O_6P_3F_{24}$) were applied. During ongoing experiments, the hardware of the instrument was upgraded (exchange of RF generator, installation of new ion cooler / higher energy collisional-dissociation collision cell, new micro-channel plate detector). UV profile was recorded from 190–800 nm with 1.2 nm resolution.

6.1.7 LC-CID MS/MS analysis

For significantly different features between control and transgenic *Lr34* groups, targeted CID-MS/MS product ion scan experiments were performed using samples with high content of the corresponding feature. The same basic parameters described above were applied, with the following adjustments; *m/z* range 50–1000, spectral rate 10 Hz, precursor

isolation width 8. The collision cell was operated with 0.8 ml nitrogen as collision gas. Specific parameters depending on the feature of interest were selected independently: retention time window, precursor ion mass, collision energy ramp (e.g. 10–35 eV). Acquisitions in both ionization modes were conducted with dead volume mass calibration using sodium formiate clusters.

6.1.8 Data processing

Internal mass calibration was done in *Bruker DataAnalysis* 4.1 software from 0.1–0.15 min with HPC mode. Calibrated data files were converted to .mzXML format with *Bruker CompassXport* 3.0.7 software. Signals of selected features were normalized with internal standards ampicillin (RT 2.0 min, *m/z* $[M+H]^+$ 350.1169) and camphor sulfonic acid (RT 1.9 min, *m/z* $[M–H]^–$ 231.0685).

XCMS Online processing

XCMS Online 2.00.00 (https://xcmsonline.scripps.edu/) was processed with standard parameters suggested for UPLC / Bruker Q-TOF.
General: RT format minutes, polarity negative, Feature detection: method centWave, 10 ppm, min. peak width 5, max. peak width 20, RT correction: method orbiwarp, profStep 1, Alignment: mzwid 0.015, minfrac 0.5, bw 5, minsamp 1, max 100, Statistics: Welch's *t*-test, no paired test performed, 1.5 fold-change threshold, *p*-value threshold (highly significant features) 0.001, *p*-value threshold (significant feature) 0.05, Annotation: search for isotopes + adducts, *m/z* abs error 0.015, 5 ppm, Identification: 10 ppm, adducts, Visualization: EIC width 200 s, Miscellaneous: no correction of mass calibration gaps, no bypass file sanity check.

Preprocessing for other statistical methods

For the statistical analysis based on *MetaboLyzer*, all data files were preprocessed in a single folder. For the statistical analysis based on *Feature filtering*, data files were separated into individual group for control and transgenic plants. Preprocessing was performed by *XCMS* in *R* 2.15.3 with the following parameters for peak picking: xcmsSet algorithm, files NULL, method centWave, nSlaves 5, peakwidth c(5,60), prefilter c(2.50), for grouping: group algorithm, minfrac 0.25, for retention time correction: rector, with further grouping as above.

***MetaboLyzer* processing**

MetaboLyzer was run on *an Oracle VM VirtualBox, Linux Debian* 3.2.0-4-amd64 on a *Mac OS X* 10.7.5. *MetaboLyzer* v7.2.0 was run in GNOME terminal 3.4.1.1 with the following parameters: Ion presence percentage cutoff 0.99, data transform log, data normalization standard Gaussian, no IQR outlier removal, *p*-value 0.05, with putative metabolite identification, *KEGG* human, *BioCyc MetaCyc*, molecular species negative/positive depending on ionization, adducts all, molecular weight tolerance 20 ppm, Test-filter Mann-Whitney U-test, permutation testing with 1000 replicates.

Processing based on *Feature filtering*

The *Feature filtering* method is based on [202], conducted in *R*. For the preprocessed data, a matrix of peak values for each group was extracted (parameter: maxo), mean intensities were calculated for each group, and filtered with threshold value for MS/MS (1000). The intensity distribution was plotted. Filtered intensities were normalized across samples to max=1, and visualized in a PCA. Than, the percentage of features where the coefficient of variation of samples was above the coefficient of variation of the quality controls was defined, and combined with the intensity filtering. A new dataset was created without quality controls. Multivariate exploratory data analysis was performed, and again visualized in PCA. Data were adjusted for multiple testing using false discovery rate, and the fold change for the transgenic line versus the control was calculated.

6.1.9 Calculation of unknown molecular formula

The molecular formula was calculated using the *SmartFormula3D* algorithm (*DataAnalysis* 4.1 and 4.2, *Bruker Daltonics*), which restricts the number of formulas by considering fragments as subset of the respective formula, with the following parameters: maximum elemental composition for (–)-ESI $C_aH_bN_cO_dPS$, for (+)-ESI $C_aH_bN_cO_dNa_eK_f$ PS with $1 \leq b/a \leq 3$; e=0 or 1; f=0 or 1; a, b, c, and d not limited. Rings plus double bonds values from –0.5 to 40, the nitrogen rule and ions of even electron configuration were considered.
Limits for mass accuracy deviation were estimated to be below 1.5 ppm using *m/z* of internal standards and lock masses, and a tolerance of 5 ppm was used.

6.1.10 Hydrolysis

Hydrolysis apparatus was operated under argon inert gas (Figure 40) at reflux.

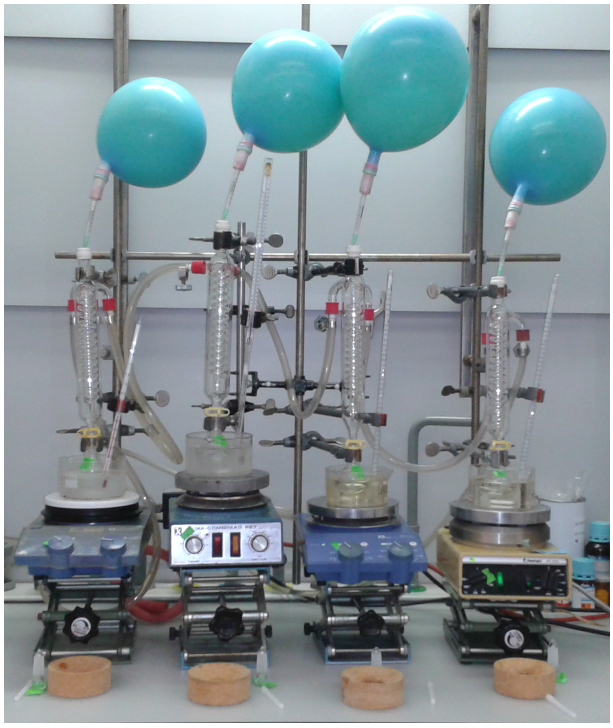


Figure 40: Hydrolysis apparatus under argon.

The following samples were processed (Table 14): reference solutions of apigenin, genkwanin, naringenin, sakuranetin, vitexin and hesperetin (Ref), sister line plant material spiked with reference solutions (Sib+Ref), sister line plant material (Sib) and BG9 plant material (BG9).

Table 14: Composition of samples for hydrolysis of plant material

	Ref	Sib+Ref	Sib	BG9
Methanol, degased [ml]	2.5	2.5	2.5	2.5
Ultrapure water, degased [ml]	1.9	1.9	1.9	1.9
Reference stock solutions (5 µg ml^{-1}) [µl], each	2.5	2.5	-	-
Plant powder [mg]	-	50	50	50

To test the effects of the different steps, aliquots were taken at start, after addition of either ascorbic acid or hydrochloric acid, after addition of both ascorbic acid and hydrochloric acid, after hydrolysis and after filtration with two different filtration units.

For hydrolysis, 10 µl ascorbic acid solution (0.2 mg ml^{-1} in ultrapure water), and 0.59 ml of HCl were added to samples, shaken and the mixture was heated to 80 ± 5 °C for 2 h at reflux. The hydrolyzed sample was diluted 1:1 with MeOH for filtration. Filtration units: *Filtropur S*, polyethersulfone (PES), 0.45 µm sterile, *Sarstedt or Puradisc 30 Syringe Filter*, cellulose acetate (CA), 0.45 µm sterile, *Whatman GE Healthcare*. Filtrates were collected in LC-MS vials and stored in autosampler at 10 °C upon analysis.

6.1.11 LC-MS analysis of hydrolysates

On an *Acquity UPLC* (Waters) controlled by *HyStar* chromatography software (*Bruker Daltonics*), separation was performed on a reversed phase analytical column *Acquity* BEH C18 (1.7 mm, 1 x 50 mm, guard column 2.1 x 5 mm, Waters) maintained at 30 °C. Acidified water / MeCN solvents containing 0.1% [*v/v*] formic acid were applied with the following linear binary gradient: 0–7 min from 3% to 100% MeCN, wash for 3 min at 100% MeCN, reequilibration 2 min at 3% MeCN at a flow rate of 0.2 ml min^{-1}. Sample volume of 2 µl was injected with needle overfill flush injection mode, with 300 µl strong (water / MeOH 1:1 [*v/v*]) and 500 µl weak needle wash (water / MeCN 95:5 [*v/v*]). After injection of aliquots hydrolysis and filtrates, blank MeOH was injected to ensure no carry-over. Mass spectra were recorded on a *maXis Bruker* QTOF instrument in (+)-ESI mode, over a *m/z* range 50–1200 at 2 Hz acquisition rate, with the following parameters: 4000 V capillary voltage, -500 V endplate offset, with a N_2 nebulizer pressure of 2.0 bar and dry gas flow of 9 l min^{-1} at 200 °C, transfer funnel RF of 200 Vpp, multipol RF of 200 Vpp, quadrupole ion energy of 3.0 eV, collision cell collision energy of 10.0 eV, collision RF 300 Vpp, transfer time of 100 µs, Ion cooler RF 100 Vpp, and 5 µs prepulse storage. Dead volume mass calibration was performed with sodium formiate clusters.

Table 15: MS-detection of reference compounds for hydrolysis

Compound	Formula	*m/z* $[M+H]^+$	RT [min]
Apigenin **12**	$C_{15}H_{10}O_5$	271.0612	3.42
Genkwanin **13**	$C_{16}H_{12}O_5$	285.0712	4.31
Naringenin	$C_{15}H_{12}O_5$	273.0751	3.36
Sakuranetin	$C_{16}H_{14}O_5$	287.0913	4.20
Vitexin	$C_{21}H_{20}O_{10}$	433.1129	2.42
Hesperidin	$C_{28}H_{34}O_{15}$	611.1971	2.76
Hesperetin	$C_{16}H_{14}O_6$	303.0863	3.53

6.2 Metabolomic profiling of *Lr34* resistance in barley, rice and wheat by LC- and GC-MS

6.2.1 Plant material

Transgenic barley and rice lines were produced by Agrobacterium-mediated transformation. For barley and rice, each a low- and high-expression line were used (Table 16).

Table 16: Transgenic barley and rice lines with low- and high expression of *Lr34*; germin-like protein GER4 (GLP) [203], references a [25] and b [26].

Species	*Lr34* line	*Lr34* expression	Promotor	Plasmid	Effects of low vs high expression line			Reference
					Expression level	LTN phenotype	Resistance	
Barley	GLP	low	Ger4c	*p6UGLP:Lr34res*	10-20x lower than in line BG8	less than line BG8	less than line BG8	-
Barley	BG8	high	native *Lr34*	*p6U:gLr34res*	-	-	-	a
Rice	8	low	native *Lr34*	*p6U:gLr34res*	4-10x lower than in line 19 (seedling stage)	only in adult plants	same as line 19	b
Rice	19	high	native *Lr34*	*pWBVec8:gLr34res*	-	already strong at seedling stage	-	b

An overview of growth conditions and timing is listed in Table 17.

Table 17: Overview of growth conditions and timing of individual batches; days post infection (dpi), hours post infection (hpi).

Species	Condition	Growth			Treatment		Harvest		
		Medium	Growth condition	Batch names	Age	Infection	Age	Leaf no.	Time after infection
Barley	Hydroponics	hydroponics	phytotron	BH1 / BH2	-	-	14 / 14 d	3	-
Barley	Mock infected	soil	phytotron	BM1 / BM2	15 / 17 d	without pathogen (mock)	23 / 28 d	4	8 / 11 dpi
Barley	Pathogen infected	soil	phytotron	BI1 / BI2	15 / 17 d	*Puccina hordei*	23 / 28 d	4	8 / 11 dpi
Rice	Hydroponics	hydroponics	phytotron	RH1 / RH2	-	-	20 / 20 d	5	-
Rice	Mock infected	soil	greenhouse	RM1 / RM2	22 / 22 d	without pathogen (mock)	23 / 23 d	4	28 / 28 hpi
Rice	Pathogen infected	soil	greenhouse	RI1 / RI2	22 / 22 d	*Magnaporthe griseae*	23 / 23 d	4	28 / 28 hpi
Wheat	Non-infected	soil	field	WN1	-	-	9 m	last	-
Wheat	Pathogen-infected	soil	field	WI1-WI4	7 m	*Puccina triticina*	9 m	last	56 dpi

Barley

The same seed batches of barley (*Hordeum vulgare)* c.v. Golden Promise were used for all growth conditions. As controls, sister lines with identical genetic background lacking *Lr34res* were used. Detailed growth dates are listed in Supplementary table 9.

For hydroponic growth, seeds were surface sterilized with sterilization solution (sodium hypochlorite 1%, Triton-X-100 0.03%) for 20 min and stratified for 3 d at 4 °C. Seeds were pre-germinated vertically in square petri plates (*Greiner*, Frickenhausen, Germany) for 5 d in phytotron. Young seedlings were transplanted into a hydroponic system (*General Hydroponics*, Sebastopol, USA), filled with nutrient solution based on Hoagland nutrient solution: K_2SO_4 2.5 mM, $Ca(NO_3)_2$ 5 mM, KH_2PO_4 0.5 mM, $MgSO_4$ 2.5 mM, KCl 0.25 mM, H_3BO_3 2.5 µM, $MnSO_4$ 2.5 µM, $CuSO_4$ 0.5 µM, $(NH_4)Mo_7O_{24}$ 0.05 µM, $ZnCl_2$ 2.5 µM, Fe-Na-EDTA 0.25 mM, pH 6.5 by KOH 1 M. Plants were grown at light cycles of 16/8 h

light/dark in phytotron climate chambers with 450 µM mercury discharge light at 65% humidity, 24 °C day / 16 °C night temperature. Whole third leaves of 14 day old plants (counted day in light) were harvested in scintillation polyvials 20 ml (Zinsser Analytic, Frankfurt am Main, Germany) and flash frozen in liquid nitrogen.

For infection experiments, barley seeds were grown on soil (Einheitserde classicProfi Substrat, Einheitserdewerke Werkverband e.V., Sinntal-Altengronau, Germany) for approx. 15 days. At the day of infection, barley leaf rust spores (*Puccinia graminis sp hordei*, strain 1.2.1) were incubated for 2 min in a 42°C water bath and diluted in infection oil (Fluorinert Electronic Liquid FC-43, 3M, St. Paul, MN, US) to a light brown solution. Plants were sprayed with water, and spores were applied via air pressure pump. Plants were incubated for 30–60 min at ambient temperature, sprayed with water and covered with a waterproof plastic dome previously humidified with water spray. Again, plants were incubated for 30 min and sprayed with water. For the next 24 h, plants were incubated at 16 °C and 90% humidity without light in growth chamber. Plants were further incubated with the following program: 16 h with light at 20 °C, 8 h without light at 16°C, both at 70% humidity. After 3 days, the plastic dome was replaced with a permeable plastic dome. After development of intense pustules, whole 4th leaves were harvested in scintillation tubes and flash frozen in liquid nitrogen. Due to less pustule development, a longer post-infection phase was chosen for batches BM2+BI2.

Rice

Rice (*Oryza sativa japonica* cv Nipponbare) seeds originating from the same growth batches were used for all growth conditions. Detailed growth dates are listed in Supplementary table 9.

For hydroponic growth, seeds were de-husked and placed on wet tissue for germination at ambient temperature in the dark for seven days. Seedlings were transferred to a home-made hydroponic system containing the following nutrient solution: Na_2O_3Si 0.1 mM, $Ca(NO_3)_2$ 1.5 mM, KH_2PO_4 0.5 mM, $MgSO_4$ 10.5 mM, KNO_3 1.25 mM, Fe-Na-EDTA 0.1 mM, KCl 50 µM, $MnSO_4$ 10 µM, $CuSO_4$ 1.5 µM, $ZnSO_4$ 2 µM, H_3BO_3 50 µM, $(NH_4)_6Mo_7O_{24}$ 0.075 µM, pH adjusted to 5.5-6.5. Plants were grown in phytotron under standardized conditions (80% humidity, 28 °C, 12 h light/dark photoperiod, light intensity 600 µM) for 19 days. The fourth leaves were harvested in scintillation tubes and flash frozen in liquid nitrogen.

For infection experiments, the fungus *Magnaporthe grisea* (rice blast) isolate Fr13 was grown on oatmeal agar media in a petri dish (30 g l^{-1} oatmeal, 5 g l^{-1} sucrose and 16 g l^{-1}

agar). Three plugs of stock mycelium were transfer on a new petri dish to start a new culture, and grown at ambient temperature in the dark for 8 days, before being transferred to blue and with light (Philips TL-D 15W BLB) for 6 days. The blue light induces sporulation of the fungus. Spores were collected by rinsing the petri dish with sterile distilled water and raking with a spatula. Spores were filtered through two layers of gauze and suspended to a density of ~ 10^5 spore per ml, using a Thoma counting chamber. In advance to application, 0.02 % of Tween 20 was added to the spore solution.

Rice soil was prepared as a mixture of 50% [*w/w*] *Rasenerde* (Ökohum, Herbertingen, Germany) containing N: 100 mg/l, P_2O_5: 80 mg l^{-1}, K_2O: 480 mg l^{-1} and Mg: 80 mg l^{-1}, 25% [*w/w*] clay powder and 25% [*w/w*] *Nulltorf* (DV Pellet*,* Białystok, Poland) from Lithuanian peat to adjust the soil pH at around 5. Fertilizer contained 0.1% [*v/v*] of *Sequestrene Rapid* and 0.2% [*v/v*] *Wuxal Hydro* (Syngenta Agro, Dielsdorf, Switzerland) in rain water (pH 6.5). Plants were sown in rice soil and grown in greenhouse at 28 °C with 12 h photoperiod, 600 µm light intensity and 80% humidity. Every day, plants were watered with rain water, and twice a week 2.5 dl fertilizer was given to the rice tank. Four to five plants were grown per pot for 22 days. Before infection, leaves were sprayed with water in order to increase the humidity of the leaf surface. Then, 5 µL droplets of blast fungus suspension (70000 spores ml^{-1}) were placed on leaf surface (~7 droplets/leaf) to increase the spore concentration in local areas of the marked leaf to ensure good infection. The remaining spore solution was sprayed over all leaves. For mock infection, a suspension without fungal spores was used. Plants were incubated in high humidity condition (~100%) for 28 h. Harvested leaves were visually observed for infection symptoms (black spots), collected in scintillation tubes and flash frozen in liquid nitrogen. After seven days, final symptoms of fungal infection were investigated.

Wheat

Winter wheat (*Triticum aestivum L.* cv Arina) was grown on fields at Agroscope Reckenholz Zürich, Switzerland (GPS 47.430361, 8.519696). After seven months, pots wheat plants infected with Puccinia triticina were transplanted to the field. After nine month, samples were collected from different patches (small experimental fields) all over the big field; one patch of non-infected Arina control and Arina *Lr34* respectively, and each four patches of pathogen infected Arina control and Arina *Lr34*. Of each patch, flag leaves of nine replicate plants were harvested, cut and kept on dry ice. Harvesting time was 10:45–14:30 p.m. on 03.06.2014; weather was humid and sunny with a thunderstorm later in the day.

6.2.2 Sample preparation

Homogenization and weighing

Pre-chilled stainless steel balls (3 x 5 mm, 2 x 3 mm for barley and rice, 2 x 8 mm, 2 x 5 mm, 2 x 3 mm for wheat) were added to leaf samples in scintillation tubes. For infected rice with high *Lr34* expression, two leaves were pooled to get sufficient plant material. A *Cryogenic Grinder System* (Labman Automation Ltd, North Yorkshire, GB) was operated at –60 °C with around 10 % humidity and leaves were homogenized by grinding twice 90 s at 60 Hz, with an intermediate break of 60 s. Aliquots of plant powder (50 ± 2 mg) were weighed in frozen state into pre-chilled 2.0 ml safe lock tubes (Eppendorf, Hamburg, Germany) and stored at –70 °C upon extraction.

Extraction

The following chemicals were used for extraction: MeOH CAS 67-56-1, tert-butyl methyl ether (MTBE) CAS 1634-04-4, ultrapure water ULC/MS (all in ULC/MS quality, Biosolve, Dieuze, France), ampicilline CAS 69-53-4, corticosterone CAS 50-22-6, and [$^{13}C_6$]-sorbitol (Sigma–Aldrich, Taufkirchen, Germany), 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine CAS 70897-27-7 (Avanti, Alabaster, AL, US).
Extraction mixture 1 (methanol with 2 µg ml^{-1} corticosterone: MTBE with 0.67 µg ml^{-1} lipid standard 1:4 [v/v] including 0.5 µg ml^{-1} ^{13}C sorbitol and 0.25 µg ml^{-1} ampicilline) was pre-cooled for 30 min at –15 °C. Samples were kept on dry ice, and 1 ml of extraction mixture 1 was added, vortexed and stored on ice. Extracts were incubated in a *Thermomixer 5436* (Eppendorf, Hamburg, Germany) at 4 °C for 10 min, and in ultrasound bath cooled with ice for another 10 min. At ambient temperature, 500 µl of extraction mixture 2 (water : MeOH 3:1 [v/v]) were added, vortexed and centrifuged (2 min, 20817 g). Of the upper organic, green phase, aliquots of 600 µl were transferred to 1.5 ml vial for lipid analysis. For determination of chlorophyll content, absorbance of the organic phase at 665 nm was measured (dilution in MeOH to OD_{665nm} <0.5). Remaining organic phase was removed completely by aspiration with vacuum pump, and samples were again centrifuged. Aliquots of lower polar phase were transferred to 1.5 ml vials for secondary metabolite LC-MS (430 µl) and for GC-MS-analysis (150 µl) respectively. Aliquots were dried to complete dryness in centrifugal vacuum concentrators *Scanvac Scan Speed 40* and *Cold Safe* (LaboGene, Lynge, Denmark) at 30 °C, <10 mbar, and stored at –20°C upon reconstitution.

Derivatization

Before and after each sample sequence measured by GC-MS, a quality control mixture was evaluated daily. Quality control mixture was prepared by dissolving various reference compounds (Supplementary table 10), Sigma–Aldrich, St. Louis, MO, US) in appropriate solvents, combined to a total volume of 900 ml of H_2O: MeOH: $CHCl_3$ 1:2.5:1 [*v/v/v*]. Aliquots of 270 µl were dried to complete dryness, stored at –20 °C and derivative with each daily sample set with double volume (20/180 µl methoxamine solution /*N*-methyl-*N*-trimethylsilyltrifluorocetamide (MSTFA) CAS 24589-78-4, (Macherey-Nagel, Düren, Germany) -fatty acid methyl ester (FAME) mixture).

Samples stored at –20 °C were defrosted for approx. 20 min, and 10 µl methoxamine solution (40 mg *O*-methylhydroxylamine hydrochloride CAS 593-56-6 (Sigma–Aldrich, Taufkirchen, Germany) dissolved in 1 ml pyridine CAS 110-86-1 (Merck, Darmstadt, Germany)) was added via multipipette *Stream*, (Eppendorf, Hamburg, Germany). For methoxyamination, samples were incubated in heating *thermomixer HLC TM130-6* (HLC Ditabis, Pforzheim, Germany) at 30 °C for 90 min. After centrifugation (30 s, 14000 rpm/20817 g), 90 µl MSTFA-FAME (100 µl FAME mixture (Supplementary table 11) in 10 ml MSTFA) was added and incubated at 37 °C for 30 min for trimethylsilyl derivatization. For blank samples, 10 µl methoxamine solution was mixed with 90 µl MSTFA without FAME. After centrifugation (30 s, 20817 g), samples were transferred to GC-MS vials. All pipetting steps for derivatization were performed quickly to avoid extended oxygen exposure.

Hormone extraction

For extraction, the following chemicals were purchased: MeOH CAS 67-56-1, ultrapure water CAS 7732-18-5, isopropanol CAS 67-63-0, (all in ULC/MS quality, Biosolve, Dieuze, France), dichloromethane CAS 75-09-2 (puriss p.a., Sigma–Aldrich,), salicylic acid CAS 69-72-7, salicylic acid d_4 CAS 78646-17-0, jasmonic acid CAS 6894-38-8, abscisic acid CAS 14375-45-2, abscisic acid d_6 CAS 35671-08-0, all from *OlChemim* (OlChemim, Olomouc, Czech Republic).

Extraction for plant hormones was carried out as described by [186]. For rice samples (without and with pathogen), only 30 mg of plant tissue was used with subsequently lower volumes for extraction. For all other samples of rice, barley and wheat, 50 mg plant powder were used. Briefly, tissue was extracted with extraction solvent (1-propanol / H_2O / concentrated HCl (2 : 1 : 0.002 [*v/v/v*]) containing 0.1 µl ml^{-1} internal standards d_4-SA, d_6-

ABA, d_5-IAA and d_6-JA) with a ratio 1:10 [*w/v*] plant powder/extraction solvent shaking for 30 min at 4 °C. Dichloromethane was added (ratio 1:20 [*w/v*] plant powder/dichloromethane) and incubated for another 30 min at 4 °C. After centrifugation for 5 min at 4 °C, aliquots of the lower phase (900 µl / 540 µl respectively) were transferred to a fresh tube. Aliquots were dried in speedvac (Scanvac Scan Speed 40 and Cold Safe, LaboGene, Lynge, Denmark) to approximately 15 µl, and resuspended in totally 100 µl / 60 µl MeOH respectively.

6.2.3 Analytic methods

For LC-MS analyses, all chemicals were purchased in analytical grade from Sigma–Aldrich (Taufkirchen, Germany) and solvents UPLC grade from Biosolve (Dieuze, France), respectively.

LC-MS secondary metabolites

For secondary metabolite sample reconstitution, 200 µl of ultrapure water was added to the dry aliquots. After allowing to dissolve for 15 min at ambient temperature, samples were vortexed, sonicated for 5 min, centrifuged (2 min, 20817 g), and 170 µl supernatant was transferred to LC-MS vials. Of all samples analyzed at one day, a pool was prepared for quality control.

Samples were analyzed by ultra-high-performance liquid chromatography - Fourier transform mass spectrometry (UHPLC/FT-MS) on an *Acquity UPLC* (Waters, Milford MA, US) coupled to an *Thermo Exactive Orbitrap* MS1.1 (Thermo Scientific, Waltham, MA, US) mass spectrometer. Separation was performed with a reverse-phase column *Acquity UPLC HSS T3* (1.8 µm particle size, 2.1x100 mm, equipped with 2.1x5 mm precolumn, Waters) at 40°C with 3 µl injection volume. Injection needle wash consisted of ultrapure water : MeCN 9 : 1 [*v/v*] (weak wash, 500 µl), and MeCN (strong wash, 500 µl). The mobile phases for the gradient consisted of acidified ultrapure water (A) and acidified MeCN (B) with each 0.1% [*v/v*] formic acid. The gradient was run with flow rate 0.4 ml min^{-1} as follows: 1 min, 1% B; 10 min linear gradient from 1% to 40% B; 3 min linear gradient from 40% to 70% B; 2 min linear gradient from 70% B to 99% B. After washing the column for 1 min with 99% B, the column is re-equilibrated back at 1% B for 3 min (total run time 20 min).

The parameters for mass spectral acquisition were as follows: mass range *m/z* 100–1500 at enhanced resolution (25000 at 4 Hz), heated ESI source sheath gas flow rate 60 ml min^{-1}, aux gas flow 35 ml min^{-1}, spray voltage 3 kV, capillary temperature 250 °C,

heater temperature 350 °C. Parameters in positive ionization mode were capillary voltage 25 V, tube lens voltage 130 V, Skimmer voltage 25 V and respectively –40 V, 120 V and -25 V in negative mode. Spectra were recorded alternating between full-scan and all-ion scan modes. For both polarities, separate injections were performed.

LC-MS lipids

The dried lipid extracts were re-suspended in 200 µl of MeCN : isopropanol 7:3 [*v/v*]. After allowed dissolving for 15 min at ambient temperature, samples were vortexed, sonicated for 5 min, centrifuged (2 min, 20817 g), and 170 µl supernatant was transferred to LC-MS glass vials. Of all samples analyzed at one day, a pool was prepared for quality control.
The same instrument as described in LC-MS secondary metabolites was used for lipid analysis, and the method was based on [144, 204]. Separation was achieved by an *Acquity UPLC BEH C8* column (1.7 µm particle size, 2.1x100 mm, Waters) kept at 60°C, with of 2 µl injection volume. Injection needle wash consisted of ultrapure water : MeCN : isopropanol 20 : 56 : 24 [*v/v/v*] (weak wash, 500 µl), and MeCN (strong wash, 500 µl). The mobile phases were solvent A: water with 1% [*v/v*] 1 M ammonium acetate and 0.1% [*v/v*] acetic acid; and solvent B, MeCN : isopropanol (7 : 3, containing 1% 1 M CH_3COONH_4 and 0.1% acetic acid). The gradient was run with flow rate of 0.4 ml min^{-1} with the following profile: 1 min 55% B; 3 min linear gradient from 55% B to 75% B; 8 min linear gradient from 75% B to 89% B; 3 min linear gradient from 89% B to 100% B. The column was washed for 4.5 min with 100% B, and re-equilibrated for 4.5 min at 55% B (total run time 24 min). Parameters for MS acquisition were as follows: mass range *m/z* 100–1500 at enhanced resolution (25000 at 4 Hz), heated ESI source sheath gas flow rate 60 ml min^{-1}, aux gas flow 20 ml min^{-1}, spray voltage 3.5 kV, capillary temperature 250 °C, capillary voltage 27.5 V, tube lens voltage 150 V, skimmer voltage 20 V and heater temperature 350 °C. Alternating full-scan and all-ion fragmentation scan spectra were recorded, and polarity was switched from negative to positive after 12 min. An example chromatogram is shown in Supplementary figure 16.

LC-MS hormones

The UHPLC-MS method described by [205], was used to detect the plant hormones. In brief, 2 µl MeOH extract was chromatographically separated on an *Acquity UPLC HSS T3 C18* column (100 x 2.1 mm, 1.8 µm particle size) at 40°C column temperature. Mobil phases containing 0.1% formic acid in H_2O (A) and MeCN (B) were applied with flow rate 400 µl/min with the following profile: isocratic 0–1 min at 25% B, linear gradient 1–6 min

25–99% B, isocratic 6–7 min 99% B, re–equilibration at 25% B, with a total runtime of 12 min. Mass spectra were acquired over a range of *m/z* 100–700, with resolution 10000 and transfer capillary temperature of 200 °C. In preliminary experiments, solutions of non-labeled plant hormone standards were used to evaluate retention times.

GC-MS primary metabolites

Gas chromatograph *6890N Network GC System* (Agilent, Santa Clara, CA, US) was coupled to electron impact ionization at 70 V and a *Pegasus IV 5ml GC/TOF Model 614-200-600*, controlled with *Leco ChromaTOF* 3.34 (Leco, St. Joseph, MO, US). The details of the GC-EI-TOF-MS method are as follows: 1 µl was injected at 230 °C in splitless mode with helium carrier gas 2 ml min^{-1}, with 90 s at 20 ml/min purge flow and head pressure 6.3 psi. GC separation was performed on an *Agilent J&W DB-35* column, (28.981 m x 0.32 mm x 0.25 µm, Agilent, Santa Clara, CA, US). The temperature program started with 2 min isothermal mode at 85 °C, followed by a 15 °C min^{-1} ramp to 360 °C, and the final temperature was kept constant for 2 min. Mass spectrometer ion source was set to 250 °C, the filament bias current to –70 V and the mass spectrometer was operated over mass range *m/z* 85–750 with 20 Hz acquisition rate with detector voltage 1500 V.

6.2.4 Data processing

LC-MS data processing secondary metabolites and lipids

LC-MS data raw files were processed with *Genedata Refiner MS* 7.5 software (Genedata, http://www.genedata.com), parameters resembling previously reported studies [144, 164]. In a first step, chemical noise was subtracted (chromatogram smoothing with RT window 3 scans, chemical noise subtraction with RT window 51 scans, quantile 50% and intensity threshold 500, RT structure removal minimum RT length 6 scans, *m/z* structure removal of minimum *m/z* length 3 points). After intensity thresholding based on clipping of profile data on 500 intensity, chromatogram retention time alignment was performed using a pairwise alignment based tree with m/z windows of 5 points and RT windows of five scans. For curvature-based peak detection, summation window was 10 scans, minimum peak size 5 scans and maximum merge distance 4 points, peak RT splitting on maximum intensity profiling 90%, smoothing window 3 points, peak refinement performed with 80% threshold, and a consistency threshold of 1. The output of *Genedata* was a matrix of all samples with intensities, and a list with peak information (mean *m/z*, RT). Isotopes were removed from the preprocessing output tables by using the *R* package *gdps* [206] with *m/z* tolerance

0.008 and RT tolerance 0.08 min. Samples were sorted according the same order for all platforms.

Data normalization was performed using *R* software [207, 208] in three steps. First, data was normalized to pool factors, calculated from mean pools of barley+rice / or wheat of the respective day divided by the mean of all pools of the dataset. Than, data were filtered for signal intensities above two times mean intensities in blank samples, to remove impurities. Next, a factor for the internal standards was calculated based on signal intensity divided by median of the whole dataset. Pool-normalized internal standard signals for secondary metabolites were $[M+H]^+$ and $[M+HCOO]^-$ of corticosterone at RT 11.40 min, for lipids $[M+CH_3COO]^-$ of 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine at RT 10.00 min. Finally, the peak intensities were normalized to a factor of fresh weight, calculated by the sample´s individual fresh weight divided by median of all fresh weights of the dataset.

The normalized peak intensities were filtered according to occurrence, robustness and abundance. Occurrence was defined as minimum one group containing six (barley and rice) or nine (wheat) replicates to be fully occupied. For robustness, the coefficient of variation (CV) of each group was calculated for each peak, and the global mean of all CV´s had to be below 50%. For abundance filtering, a cutoff was evaluated based on intensity of peaks that were only occupied in single replicates per group, and peaks with maximum group mean below this cutoff were filtered away.

For lipid annotation, an in-house developed *R* package was applied [164], using the library compiled by [144]. Basically, the package uses previously identified markers with known retention times to calculate a retention-time index (RI). To annotate the lipids, the peak list was compared to the library based on their *m/z* and RI, with tolerances of 10 ppm and 0.1–1 min. Annotations were manually inspected by observation of the MS trace in (+)- and (–)-ESI in *Xcalibur* 3.0.63 (Thermo Fisher Scientific, Waltham MA, US).

For principal component analysis, the normalized and filtered data were log2 transformed. Data were visualized with *R* package *pcaMethods* 3.1 [209]. For heatmaps, *R* packages *gplots* 2.11.0.1 [210] was applied on log2 fold change data.

LC-MS data processing hormones

Hormone processing method was set up in *XCalibur* 2.1.0.1140 (Thermo Scientific, Waltham, MA, US), with retention time and accurate *m/z* and the following parameters: RT window 50 s, ICIS peak integration: smoothing points 5, baseline window 150, area noise factor 5, peak noise factor 10, constrain peak width: peak with 2%, tailing factor 3, ICIS peak detection: highest peak, minimum peak height S/N 3. Peak integration was manually

curated in *Quan Browser* 3.0.63 (Thermo Scientific, Waltham, MA, US). For salicylic and abscisic acid, peak areas were corrected with ratio of peak area: peak area of labeled internal standard.

GC-MS data processing primary metabolites

For GC-MS data preprocessing, raw data were converted from .peg (Pegasus Document) format into .cdf format in *Leco ChromaTOF* software optimized for Pegasus 4.44.0.0 (Leco, St. Joseph, MI, US) with the following parameters: Compute baseline, calculate area / height, export data in .cdf format, baseline offset 1, smoothing 5, use unique mass, expected peak width 3.5 for 0 s, 6.0 for 1440 s. Peak picking was run on .cdf files with *PeakFinder* tool of *TagFinder* 4.1 [145] with the following parameters: Smooth width apex finder 10, low intensity threshold 50, quantification mode peak hight, max merging time width 0.05. Peak lists were imported with threshold 200, and for barley and rice restricted mass ranges (*m/z* 85–200, 201–400, 401–750) were used due to high number of samples. Retention time indices of peaks were calculated based on the fragment at *m/z* 87 of 13 FAMES, ranging from 262320–111310 RI units. The time scan width parameter of *TagFinder* was optimized based on a set of known metabolites and set to 500. In order to use *TagFinder* for all samples, a workspace over the whole *m/z* range 85–750 was created and data processed with the following settings: time scan width 500, gliding median group count 1, min fragment intensity 1, tag mass 85–146; 150–750, sample count 6, intensity aggregation max_intensity, tag correlation method Pearson, maximum tag distance 0.2, cor significance level sig_0001, maximum IQR ratio distance 0.04, minimum number of sample pairs 6, minimum sample group pair count 0.0, tag clustering core adjacency option same_core, min core option input_value, tag output max intensity ranking 20, min cluster size 1.

For annotation, GMD library was used [211] in *Profile Builder* tool of *TagFinder*; sample intensity matrix used tag grouping according clusters, min aggregation value count 1, tag intensity normalization max_intensity, scaling to max intensity unit 100, MS library matching with hit identifier spectrum, match mode forward, max intensity unit 0, searching time window by factor, positive and negative spectra time index 0.01, time slope 1.0, time intercept 0, in time min match value 800, match in time window only, min MMP count 80.

Manual verification of metabolites was based on clusters and verified in *ChromaTOF* software 4.44.0.0 optimized for Pegasus (Leco, St. Joseph, MI, US). Based on selected quantification masses for each metabolite, peak intensities were retrieved from the tag file. Peak intensities from the tag file were normalized in *R* [207] as described for LC-MS

secondary metabolites and lipids according to pool, internal standard ($^{13}C_6$ sorbitol, RI 587733, fragment *m/z* 220) and fresh weight.

Statistical analysis

Statistical analysis was based on normalized (pool, internal standard, fresh weight), and filtered data matrices of identified primary metabolites, LC-(+)-ESI-MS and LC-(–)-ESI-MS lipids, and LC-(+)-ESI-MS and LC-(–)-ESI-MS secondary metabolite features, and was done in *R* [207].

A two-way analysis of variance (ANOVA) was performed with regards to the groups (low expression control and *Lr34*, high expression control and *Lr34*). The *p*-value was corrected for multiple testing with Benjamini-Hochberg procedure [212], and only peaks with corrected *p*-values below a significance threshold of 0.05 were considered for the post-hoc test. To evaluate which groups are causing the significant difference, a TukeyHSD post-hoc test based on the groups (control low and high expression, *Lr34* low and high expression) was done, and only peaks with post-hoc *p*-values < 0.05 were used for the Venn diagrams. Venn diagrams were created with Venny 2.1 [213] or by [214].

6.2.5 Secondary metabolite annotation

The LC-MS method used in 6.2.3 for the analysis of secondary metabolites was transferred to a *Acquity UPLC* (Waters, Baden-Dättwil, Switzerland) coupled to a *maXis* ESI-QTOF (*Bruker Daltonics*) and a *Dionex UltiMate 3000* coupled to a *QExactive* Orbitrap (*Thermo Fischer Scientific*, Waltham, MA, US), operated with *Xcalibur* 3.0.63 at the Department of Chemistry at the University of Zurich. Both instruments were run without UV detector to minimize the dead volume.

The same chromatographic column and the same wash solutions and eluents as described in 6.2.3 LC-MS method of secondary metabolites were utilized. In order to match the original retention time as closely as possible to the original data, the gradient was slightly adjusted: 0–1 min 1% B, 1–10.2 min 1–40%B, 10.2–13.2 min 40–70% B, 13.2–15.2 min 70–99 % B, 15.2–16.2min 99% B, 16.3–20 min 1% B.

The QTOF instrument was run with source parameters 500 V end plate offset, 4000 V capillary voltage, 3.0 and 3.5 bar nebulizer gas for (+) and (–)-ESI respectively, 10 l min^{-1} dry gas at 250 °C. Mass spectra were recorded over a range of *m/z* 100–1500, with 2 Hz acquisition rate, MS/MS precursors isolated over 4 m/z, with a collision energy ramp adjusted to the metabolite.

Orbitrap source parameters were set at 50 sheat gas flow rate, 15 aux gas flow rate, 3 kV spray voltage, capillary temperature 250 at (+) and 300 °C for (–)-ESI, with S-les RF level 300 and 350 for (+)- and (–)-ESI respectively. The MS method applied a full MS/ddMS2 experiment with an inclusion list defining a *m/z* of interest in a defined RT range, full MS was recorded at 70000 resolution with 5e5 AGC target, maximum IT 100 ms from *m/z* 100–1500. For the data dependent MS/MS, the following parameters were applied: 35000 resolution, AGC target 5e4, maximum IT 50 ms, Loop count 5, TopN 5, Isolation window 0.8 *m/z*, stepped NCE 25, 35, 45 eV, dynamic exclusion off.

Chemical formula of unknown secondary metabolite features were calculated as described in 6.1.9.

7 References

[1] United Nations Agricultural Development Economics (ESA), New York, US, **2015**, *World Population Prospects: The 2015 Revision*

[2] United Nations Agricultural Development Economics (ESA), Rome, Italy, **2012**, *World Agriculture towards 2030/2050: The 2012 Revision*

[3] Food and Agriculture Organization of the United Nations, Statistics Division (FAOSTAT), http://faostat3.fao.org, accessed 10.03.2016

[4] Boggess, M. V., Lippolis, J. D., Hurkman, W. J., Fagerquist, C. K., Briggs, S. P., Gomes, A. V., Righetti, P. G., Bala, K., *J. Proteomics,* **2013**

[5] Priyamvada, M., Saharan, S., Tiwari, R., *Int. J. Genet. Mol. Biol.,* **2011**, *3*, 8, 108-114

[6] Zhang, X. Y., Loyce, C., Meynard, J. M., Monod, H., *Eur. J. Agron.,* **2007**, *26*, 4, 384-393

[7] Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., Rudd, J. J., Dickman, M., Kahmann, R., Ellis, J., Foster, G. D., *Mol. Plant Pathol.,* **2012**, *13*, 4, 414-430

[8] Zhang, Y., Lubberstedt, T., Xu, M., *J Genet Genomics,* **2013**, *40*, 1, 23-35

[9] Wilson, R. A., Talbot, N. J., *Nat. Rev. Microbiol.,* **2009**, *7*, 3, 185-195

[10] Duan, G., Christian, N., Schwachtje, J., Walther, D., Ebenhöh, O., *Metabolites,* **2013**, *3*, 1-23

[11] Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., Leung, H., *Trends Genet.,* **2013**, *29*, 4, 233-240

[12] Hammond‐Kosack K.E., Kanyuka, K., *eLS,* **2007**, 1-15

[13] Grosskinsky, D. K., van der Graaff, E., Roitsch, T., *Plant Sci.,* **2012**, *195*, 54-70

[14] Balmer, D., Flors, V., Glauser, G., Mauch-Mani, B., *Front. Plant Sci.,* **2013**, *4*, 82, 1-12

[15] Ballini, E., Lauter, N., Wise, R., *Front. Plant Sci.,* **2013**, *4*, 117, 1-11

[16] Poland, J. A., Balint-Kurti, P. J., Wisser, R. J., Pratt, R. C., Nelson, R. J., *Trends Plant Sci.,* **2009**, *14*, 1, 21-29

[17] Krattinger, S. G., Lagudah, E. S., Spielmeyer, W., Singh, R. P., Huerta-Espino, J., McFadden, H., Bossolini, E., Selter, L. L., Keller, B., *Science,* **2009**, *323*, 1360-1363

[18] Fu, D., Uauy, C., Distelfeld, A., Blechl, A., Epstein, L., Chen, X., Sela, H., Fahima, T., Dubcovsky, J., *Science,* **2009**, *323*, 1357-1360

[19] Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., van Daelen, R., van der Lee, T., Diergaarde, P., Groenendijk, J., Topsch, S., Vos, P., Salamini, F., Schulze-Lefert, P., *Cell,* **1997**, *88*, 5, 695-705

[20] Herrera-Foessel, S. A., Singh, R. P., Lillemo, M., Huerta-Espino, J., Bhavani, S., Singh, S., Lan, C., Calvo-Salazar, V., Lagudah, E. S., *Theor. Appl. Genet.,* **2014**, *127*, 4, 781-789

[21] Moore, J. W., Herrera-Foessel, S., Lan, C., Schnippenkoetter, W., Ayliffe, M., Huerta-Espino, J., Lillemo, M., Viccars, L., Milne, R., Periyannan, S., Kong, X., Spielmeyer, W., Talbot, M., Bariana, H., Patrick, J. W., Dodds, P., Singh, R., Lagudah, E., *Nat. Genet.,* **2015**, *47*, 12, 1494-1498

[22] Krattinger, S. G., Lagudah, E. S., Wicker, T., Risk, J. M., Ashton, A. R., Selter, L. L., Matsumoto, T., Keller, B., *Plant J.,* **2011**, *65*, 3, 392-403

[23] Risk, J. M., Selter, L. L., Krattinger, S. G., Viccars, L. A., Richardson, T. M., Buesing, G., Herren, G., Lagudah, E. S., Keller, B., *Plant Biotechnol. J.,* **2012**, *10*, 4, 477-487

[24] Bolton, M. D., Kolmer, J. A., Xu, W. W., Garvin, D. F., *Mol Plant Microbe Interact.,* **2008**, *21*, 12, 1515-1527

[25] Risk, J. M., Selter, L. L., Chauhan, H., Krattinger, S. G., Kumlehn, J., Hensel, G., Viccars, L. A., Richardson, T. M., Buesing, G., Troller, A., Lagudah, E. S., Keller, B., *Plant Biotechnol. J.,* **2013**, *11*, 7, 847-854

[26] Krattinger, S. G., Sucher, J., Selter, L. L., Chauhan, H., Zhou, B., Tang, M., Upadhyaya, N. M., Mieulet, D., Guiderdoni, E., Weidenbach, D., Schaffrath, U., Lagudah, E. S., Keller, B., *Plant Biotechnol. J.,* **2015**, *14*, 5, 1261-1268

[27] Chauhan, H., Boni, R., Bucher, R., Kuhn, B., Buchmann, G., Sucher, J., Selter, L. L., Hensel, G., Kumlehn, J., Bigler, L., Glauser, G., Wicker, T., Krattinger, S. G., Keller, B., *Plant J.,* **2015**, *84*, 1, 202-215

[28] Spielmeyer, W., Singh, R. P., McFadden, H., Wellings, C. R., Huerta-Espino, J., Kong, X., Appels, R., Lagudah, E. S., *Theor. Appl. Genet.,* **2008**, *116*, 4, 481-490

[29] Singh, R. P., *Crop Sci.,* **1992**, *32*, 4, 874-878

[30] Singh, R. P., Huerta-Espino, J., *Crop Sci.,* **1997**, *37*, 390-395

[31] Huckelhoven, R., Fodor, J., Preis, C., Kogel, K. H., *Plant Physiol.,* **1999**, *119*, 4, 1251-1260

[32] Kretzschmar, T., Burla, B., Lee, Y., Martinoia, E., Nagy, R., *Essays Biochem.,* **2011**, *50*, 1, 145-160

[33] Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., Martinoia, E., *Arabidopsis Book,* **2011**, *9*, e0153, 1-25

[34] Andolfo, G., Ruocco, M., Di Donato, A., Frusciante, L., Lorito, M., Scala, F., Ercolano, M. R., *BMC Plant Biol.,* **2015**, *15*, 51, 1-15

[35] Schauer, N., Fernie, A. R., *Trends Plant Sci.,* **2006**, *11*, 10, 508-516

[36] Simo, C., Ibanez, C., Valdes, A., Cifuentes, A., Garcia-Canas, V., *Int. J. Mol. Sci.,* **2014**, *15*, 10, 18941-18966

[37] Valluru, R., Reynolds, M. P., Salse, J., *Theor. Appl. Genet.,* **2014**, *127*, 7, 1463-1489

[38] Gauthier, L., Atanasova-Penichon, V., Chereau, S., Richard-Forget, F., *Int. J. Mol. Sci.,* **2015**, *16*, 10, 24839-24872

[39] Nafisi, M., Fimognari, L., Sakuragi, Y., *Phytochemistry,* **2015**, *112*, 63-71

[40] Parker, D., Beckmann, M., Zubair, H., Enot, D. P., Caracuel-Rios, Z., Overy, D. P., Snowdon, S., Talbot, N. J., Draper, J., *Plant J.,* **2009**, *59*, 5, 723-737

[41] Jones, O. A. H., Maguire, M. L., Griffin, J. L., Jung, Y.-H., Shibato, J., Rakwal, R., Agrawal, G. K., Jwa, N.-S., *Eur. J. Plant Pathol.,* **2010**, *129*, 4, 539-554

[42] Allwood, J. W., Heald, J., Lloyd, A. J., Goodacre, R., Mur, L. A. J., in *Plant Metabolomics: Methods and Protocols, Vol. 860* (Eds.: Hardy, N. W., Hall, R. D.), **2012**, pp. 31-49.

[43] Choi, D., Bostock, R. M., Avdiushko, S., Hildebrand, D. F., *Proc. Natl. Acad. Sci. U.S.A.,* **1994**, *91*, 6, 2329-2333

[44] Hazehzarghani, H., Paranidharan, V., Abu-Nada, Y., Kushalappa, A. C., Mamer, O., Somers, D., *Can. J. Plant Sci.,* **2008**, *88*, 789–797

[45] Gunnaiah, R., Kushalappa, A. C., *Plant Physiol. Biochem.,* **2014**, *83C*, 40-50

[46] Gunnaiah, R., Kushalappa, A. C., Duggavathi, R., Fox, S., Somers, D. J., *PLoS One,* **2012**, *7*, 7, e40695

[47] Palmer, L. J., Dias, D. A., Boughton, B., Roessner, U., Graham, R. D., Stangoulis, J. C. R., *Plant Methods,* **2014**, *10*, 27, 1-9

[48] Matthews, S. B., Santra, M., Mensack, M. M., Wolfe, P., Byrne, P. F., Thompson, H. J., *PLoS One,* **2012**, *7*, 8, e44179

[49] Dinelli, G., Segura-Carretero, A., Di Silvestro, R., Marotti, I., Arraez-Roman, D., Benedettelli, S., Ghiselli, L., Fernadez-Gutierrez, A., *J. Chromatogr. A,* **2011**, *1218*, 42, 7670-7681

[50] Ioset, J. R., Urbaniak, B., Ndjoko-Ioset, K., Wirth, J., Martin, F., Gruissem, W., Hostettmann, K., Sautter, C., *Plant Mol. Biol.,* **2007**, *65*, 5, 645-654

[51] Stamova, B. S., Roessner, U., Suren, S., Laudencia-Chingcuanco, D., Bacic, A., Beckles, D. M., *Metabolomics,* **2008**, *5*, 2, 239-252

[52] Hill, C. B., Taylor, J. D., Edwards, J., Mather, D., Bacic, A., Langridge, P., Roessner, U., *Plant Physiol.,* **2013**, *162*, 3, 1266-1281

[53] de Leonardis, A. M., Fragasso, M., Beleggia, R., Ficco, D. B., de Vita, P., Mastrangelo, A. M., *Int. J. Mol. Sci.,* **2015**, *16*, 12, 30382-30404

[54] Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., Roessner, U., *Molecular plant,* **2012**, *5*, 2, 418-429

[55] Crosatti, C., Quansah, L., Mare, C., Giusti, L., Roncaglia, E., Atienza, S. G., Cattivelli, L., Fait, A., *BMC Genomics,* **2013**, *14*, 868, 1-21

[56] Beleggia, R., Platani, C., Nigro, F., De Vita, P., Cattivelli, L., Papa, R., *J. Cereal Sci.,* **2013**, *57*, 2, 183-192

[57] Hanhineva, K., Rogachev, I., Aura, A. M., Aharoni, A., Poutanen, K., Mykkanen, H., *J. Agric. Food Chem.,* **2011**, *59*, 3, 921-927

[58] Voll, L. M., Horst, R. J., Voitsik, A. M., Zajic, D., Samans, B., Pons-Kuhnemann, J., Doehlemann, G., Munch, S., Wahl, R., Molitor, A., Hofmann, J., Schmiedl, A., Waller, F., Deising, H. B., Kahmann, R., Kamper, J., Kogel, K. H., Sonnewald, U., *Front. Plant Sci.,* **2011**, *2*, 39, 1-17

[59] Molitor, A., Zajic, D., Voll, L. M., Pons-Kühnemann, J. r., Samans, B., Kogel, K.-H., Waller, F., *Mol Plant Microbe Interact.,* **2011**, *24*, 12, 1427-1439

[60] Mikkelsen, B. L., Olsen, C. E., Lyngkjaer, M. F., *Phytochemistry,* **2015**, *118*, 162-173

[61] Kumaraswamy, K. G., Kushalappa, A. C., Choo, T. M., Dion, Y., Rioux, S., *J. Chem. Ecol.,* **2011**, *37*, 8, 846-856

[62] Kumaraswamy, G. K., Kushalappa, A. C., Choo, T. M., Dion, Y., Rioux, S., *Plant Pathol.,* **2012**, *61*, 509-521

[63] Kumaraswamy, G. K., Bollina, V., Kushalappa, A. C., Choo, T. M., Dion, Y., Rioux, S., Mamer, O., Faubert, D., *Eur. J. Plant Pathol.,* **2011**, *130*, 1, 29-43

[64] Cajka, T., Fiehn, O., *Trends Analyt. Chem.,* **2014**, *61*, 192-206

[65] Bollina, V., Kushalappa, A. C., Choo, T. M., Dion, Y., Rioux, S., *Plant Mol. Biol.,* **2011**, *77*, 4-5, 355-370

[66] Bollina, V., Kumaraswamy, G. K., Kushalappa, A. C., Choo, T. M., Dion, Y., Rioux, S., Faubert, D., Hamzehzarghani, H., *Mol. Plant Pathol.,* **2010**, no-no

[67] Perkowski, J., Stuper, K., Buśko, M., Góral, T., Kaczmarek, A., Jeleń, H., *J. Cereal Sci.,* **2012**, *56*, 3, 544-551

[68] Abidi, I., Mansouri, S., Radhouane, L., Ksouri, R., El Felah, M., Bouzid, S., *IJAIR,* **2015**, *3*, 5, 1417-1423

[69] Jacobs, A., Lunde, C., Bacic, A., Tester, M., Roessner, U., *Metabolomics,* **2007**, *3*, 3, 307-317

[70] Kogel, K. H., Voll, L. M., Schafer, P., Jansen, C., Wu, Y., Langen, G., Imani, J., Hofmann, J., Schmiedl, A., Sonnewald, S., von Wettstein, D., Cook, R. J., Sonnewald, U., *Proc. Natl. Acad. Sci. U.S.A.,* **2010**, *107*, 14, 6198-6203

[71] Sicher, R. C., Timlin, D., Bailey, B., *J. Plant Physiol.,* **2012**, *169*, 7, 686-695

[72] Marinova, K., Kleinschmidt, K., Weissenbock, G., Klein, M., *Plant Physiol.,* **2007**, *144*, 1, 432-444

[73] Hao, M., Beta, T., *J. Sci. Food Agric.,* **2012**, *92*, 10, 2062-2068

[74] Piasecka, A., Sawikowska, A., Krajewski, P., Kachlicki, P., *J. Mass Spectrom.,* **2015**, *50*, 3, 513-532

[75] Tohge, T., Ramos, M. S., Nunes-Nesi, A., Mutwil, M., Giavalisco, P., Steinhauser, D., Schellenberg, M., Willmitzer, L., Persson, S., Martinoia, E., Fernie, A. R., *Plant Physiol.,* **2011**, *157*, 3, 1469-1482

[76] Kovacik, J., Klejdus, B., Babula, P., Jarosova, M., *J. Plant Physiol.,* **2013**, *171*, 3-4, 260-268

[77] Huang, C. Y., Roessner, U., Eickmeier, I., Genc, Y., Callahan, D. L., Shirley, N., Langridge, P., Bacic, A., *Plant Cell Physiol.,* **2008**, *49*, 5, 691-703

[78] Widodo, H., Patterson, J. H., Newbigin, E., Tester, M., Bacic, A., Roessner, U., *J. Exp. Bot.,* **2009**, *60*, 14, 4089-4103

[79] Wu, D., Cai, S., Chen, M., Ye, L., Chen, Z., Zhang, H., Dai, F., Wu, F., Zhang, G., *PLoS One,* **2013**, *8*, 1, e55431

[80] Roessner, U., Patterson, J. H., Forbes, M. G., Fincher, G. B., Langridge, P., Bacic, A., *Plant Physiol.,* **2006**, *142*, 3, 1087-1101

[81] Khakimov, B., Jespersen, B., Engelsen, S., *Foods,* **2014**, *3*, 4, 569-585

[82] Sana, T. R., Fischer, S., Wohlgemuth, G., Katrekar, A., Jung, K. H., Ronald, P. C., Fiehn, O., *Metabolomics,* **2010**, *6*, 3, 451-465

[83] Alamgir, K. M., Hojo, Y., Christeller, J. T., Fukumoto, K., Isshiki, R., Shinya, T., Baldwin, I. T., Galis, I., *Plant Cell Environ.,* **2016**, *39*, 2, 453-466

[84] Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J., Ebana, K., Yano, M., Saito, K., *Plant J.,* **2015**, *81*, 1, 13-23

[85] Dong, X., Chen, W., Wang, W., Zhang, H., Liu, X., Luo, J., *J. Integr. Plant Biol.,* **2014**, *56*, 9, 876-886

[86] Hu, C., Tohge, T., Chan, S. A., Song, Y., Rao, J., Cui, B., Lin, H., Wang, L., Fernie, A. R., Zhang, D., Shi, J., *Sci. Rep.,* **2016**, *6*, 20942

[87] Hu, C., Shi, J., Quan, S., Cui, B., Kleessen, S., Nikoloski, Z., Tohge, T., Alexander, D., Guo, L., Lin, H., Wang, J., Cui, X., Rao, J., Luo, Q., Zhao, X., Fernie, A. R., Zhang, D., *Sci. Rep.,* **2014**, *4*, 5067, 1-10

[88] Coneva, V., Simopoulos, C., Casaretto, J. A., El-keramy, A., Guevera, D. R., Cohn, J., Zhu, T., Guo, L., Alexander, D. C., Bi, Y.-M., McNicholas, P. D., Rothstein, S. J., *BMC Genomics,* **2014**, *15*, 1056, 1-14

[89] Kim, G. R., Jung, E. S., Lee, S., Lim, S. H., Ha, S. H., Lee, C. H., *Molecules,* **2014**, *19*, 10, 15673-15686

[90] Liu, D., Ford, K. L., Roessner, U., Natera, S., Cassin, A. M., Patterson, J. H., Bacic, A., *Proteomics,* **2013**, *13*, 12-13, 2046-2062

[91] Tohge, T., Mettler, T., Arrivault, S., Carroll, A. J., Stitt, M., Fernie, A. R., *Front. Plant Sci.,* **2011**, *2*, 61

[92] Dubey, S., Misra, P., Dwivedi, S., Chatterjee, S., Bag, S. K., Mantri, S., Asif, M. H., Rai, A., Kumar, S., Shri, M., Tripathi, P., Tripathi, R. D., Trivedi, P. K., Chakrabarty, D., Tuli, R., *BMC Genomics,* **2010**, *11*, 648

[93] Zhou, J., Zhang, L., Li, X., Chang, Y., Gu, Q., Lu, X., Zhu, Z., Xu, G., *Metabolomics,* **2011**, *8*, 4, 529-539

[94] Kusano, M., Fukushima, A., Kobayashi, M., Hayashi, N., Jonsson, P., Moritz, T., Ebana, K., Saito, K., *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.,* **2007**, *855*, 1, 71-79

[95] Jiao, Z., Si, X. X., Li, G. K., Zhang, Z. M., Xu, X. P., *J. Agric. Food Chem.,* **2010**, *58*, 3, 1746-1754

[96] Vigani, G., Bashir, K., Ishimaru, Y., Lehmann, M., Casiraghi, F. M., Nakanishi, H., Seki, M., Geigenberger, P., Zocchi, G., Nishizawa, N. K., *J. Exp. Bot.,* **2015**

[97] Zhao, X., Wang, W., Zhang, F., Deng, J., Li, Z., Fu, B., *PLoS One,* **2014**, *9*, 9, e108020

[98] Degenkolbe, T., Do, P. T., Kopka, J., Zuther, E., Hincha, D. K., Köhl, K. I., *PLoS One,* **2013**, *8*, 5, e63637 63631-63614

[99] Do, P. T., Degenkolbe, T., Erban, A., Heyer, A. G., Kopka, J., Kohl, K. I., Hincha, D. K., Zuther, E., *PLoS One,* **2013**, *8*, 4, e60325

[100] Krattinger, S. G., Jordan, D. R., Mace, E. S., Raghavan, C., Luo, M. C., Keller, B., Lagudah, E. S., *Theor. Appl. Genet.,* **2013**, *126*, 3, 663-672

[101] De Vos, R. C. H., Moco, S., Lommen, A., Keurentjes, J. J. B., Bino, R. J., Hall, R. D., *Nat. Protoc.,* **2007**, *2*, 4, 778-791

[102] Benton, H. P., Ivanisevic, J., Mahieu, N. G., Kurczy, M. E., Johnson, C. H., Franco, L., Rinehart, D., Valentine, E., Gowda, H., Ubhi, B. K., Tautenhahn, R., Gieschen, A., Fields, M. W., Patti, G. J., Siuzdak, G., *Anal. Chem.,* **2015**, *87*, 2, 884-891

[103] Gowda, H., Ivanisevic, J., Johnson, C. H., Kurczy, M. E., Benton, H. P., Rinehart, D., Nguyen, T., Ray, J., Kuehl, J., Arevalo, B., Westenskow, P. D., Wang, J., Arkin, A. P., Deutschbauer, A. M., Patti, G. J., Siuzdak, G., *Anal. Chem.,* **2014**, *86*, 14, 6931-6939

[104] Kind, T., Fiehn, O., *BMC Bioinformatics,* **2007**, *8*, 105, 1-20

[105] Horai, H.*, et al.*, *J. Mass Spectrom.,* **2010**, *45*, 7, 703-714

[106] Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., Akiyama, K., Sakurai, T., Matsuda, F., Aoki, T., Hirai, M. Y., Saito, K., *Phytochemistry,* **2012**, *82*, 38-45

[107] Nielsen, K. A., Hrmova, M., Nielsen, J. N., Forslund, K., Ebert, S., Olsen, C. E., Fincher, G. B., Moller, B. L., *Planta,* **2006**, *223*, 5, 1010-1023

[108] Jensen, N. B., Zagrobelny, M., Hjerno, K., Olsen, C. E., Houghton-Larsen, J., Borch, J., Moller, B. L., Bak, S., *Nat. Commun.,* **2011**, *2*, 273

[109] Gerlich, M., Neumann, S., *J. Mass Spectrom.,* **2013**, *48*, 3, 291-298

[110] Nielsen, K. A., Olsen, C. E., Pontoppidan, K., Moller, B. L., *Plant Physiol.,* **2002**, *129*, 3, 1066-1075

[111] Kremer, R. J., Moncef, B.-H., *Allelopathy J,* **2009**, *24*, 2, 225-242

[112] Li, B., Bjarnholt, N., Hansen, S. H., Janfelt, C., *J. Mass Spectrom.,* **2011**, *46*, 12, 1241-1246

[113] Erb, N., Zinsmeister, H. D., Lehmann, G., Nahrstedt, A., *Phytochemistry,* **1979**, *18*, 1515-1517

[114] Pourmohseni, H., Ibenthal, W. D., Machinek, R., Remberg, G., Wray, V., *Phytochemistry,* **1993**, *33*, 2, 295-297

[115] Gleadow, R. M., Moller, B. L., *Annu. Rev. Plant Biol.,* **2014**, *65*, 155-185

[116] Neilson, E. H., Goodger, J. Q., Woodrow, I. E., Moller, B. L., *Trends Plant Sci.,* **2013**, *18*, 5, 250-258

[117] Kristensen, B. K., Burhenne, K., Rasmussen, S. K., *Phytochem. Rev.,* **2004**, *3*, 127-140

[118] Gorzolka, K., Bednarz, H., Niehaus, K., *Planta,* **2014**, *239*, 6, 1321-1335

[119] Stoessl, A., Unwin, C., *Can. J. Bot./Rev. Can. Bot.,* **1970**, *48*, 465–470

[120] Smith, T. A., Best, G. R., *Phytochemistry,* **1978**, *17*, 1093–1098

[121] Kageyama, N., Inui, T., Fukami, H., Komura, H., *J. A. Soc. Brew. Chem.,* **2011**, *69*, 4, 255-259

[122] Kohyama, N., Ono, H., *J. Agric. Food Chem.,* **2013**, *61*, 5, 1112-1116

[123] Pihlava, J.-M., *J. Cereal Sci.,* **2014**, *60*, 3, 645-652

[124] Yamaji, N., Yokoo, Y., Iwashita, T., Nemoto, A., Koike, M., Suwa, Y., Wakimoto, T., Tsuji, K., Nukaya, H., *Alcohol. Clin. Exp. Res.,* **2007**, *31*, 1 Suppl, S9-14

[125] Nomura, T., Ishizuka, A., Kishida, K., Islam, A. K., Endo, T. R., Iwamura, H., Ishihara, A., *Genes Genet. Syst.,* **2007**, *82*, 6, 455-464

[126] von Ropenack, E., Parr, A., Schulze-Lefert, P., *J. Biol. Chem.,* **1998**, *273*, 15, 9013-9022

[127] Nomura, T., Horikoshi, R., Tebayashi, S., Ishihara, A., Endo, T. R., Iwamura, H., *Genes Genet. Syst.,* **1999**, *74*, 99-103

[128] Burhenne, K., Kristensen, B. K., Rasmussen, S. K., *J. Biol. Chem.,* **2003**, *278*, 16, 13919-13927

[129] Christensen, A. B., Gregersen, P. L., Olsen, C. E., Collinge, D. B., *Plant Mol. Biol.,* **1998**, *36*, 219-227

[130] Wollenweber, E., Dietz, V. H., *Phytochemistry,* **1981**, *20*, 5, 869-932

[131] Brazier-Hicks, M., Evans, K. M., Gershater, M. C., Puschmann, H., Steel, P. G., Edwards, R., *J. Biol. Chem.,* **2009**, *284*, 27, 17926-17934

[132] Abrankó, L., García-Reyes, J. F., Molina-Díaz, A., *J. Mass Spectrom.,* **2011**, *46*, 5, 478-488

[133] Parejo, I., Jauregui, O., Viladomat, F., Bastida, J., Codina, C., *Rapid Commun. Mass Spectrom.,* **2004**, *18*, 23, 2801-2810

[134] Nuutila, A. M., Kammiovirta, K., Oksman-Caldentey, K.-M., *Food Chem.,* **2002**, *76*, 519-525

[135] Francescato, L. N., Debenedetti, S. L., Schwanz, T. G., Bassani, V. L., Henriques, A. T., *Talanta,* **2013**, *105*, 192-203

[136] Ballhorn, D. J., Elias, J. D., *Ann. Bot.,* **2014**, *114*, 2, 357-366

[137] Moller, B. L., *Curr. Opin. Plant Biol.,* **2010**, *13*, 3, 338-347

[138] Miller, R. E., Gleadow, R. M., Cavagnaro, T. R., *Plant Cell Environ.,* **2014**, *37*, 4, 929-942

[139] Batchu, A. K., Zimmermann, D., Schulze-Lefert, P., Koprek, T., *Genetica,* **2006**, *127*, 1-3, 87-99

[140] Bird, C. R., Smith, T. A., *Ann. Bot.,* **1984**, *53*, 4, 483-488

[141] Ogura, Y., Ishihara, A., Iwamura, H., *Z. Naturforsch. C,* **2001**, *56*, 3-4, 193-202

[142] Rojas, C. M., Senthil-Kumar, M., Tzin, V., Mysore, K. S., *Front. Plant Sci.,* **2014**, *5*, 17, 1-12

[143] Ribot, C., Hirsch, J., Balzergue, S., Tharreau, D., Notteghem, J. L., Lebrun, M. H., Morel, J. B., *J. Plant Physiol.,* **2008**, *165*, 1, 114-124

[144] Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H.-M., Hesse, H., Segu, S., Hummel, J., Kohl, K., Willmitzer, L., *Plant J.,* **2011**, *68*, 2, 364-376

[145] Luedemann, A., Strassburg, K., Erban, A., Kopka, J., *Bioinformatics,* **2008**, *24*, 5, 732-737

[146] Kopka, J., Schauer, N., Krueger, S., Birkemeyer, C., Usadel, B., Bergmuller, E., Dormann, P., Weckwerth, W., Gibon, Y., Stitt, M., Willmitzer, L., Fernie, A. R., Steinhauser, D., *Bioinformatics,* **2005**, *21*, 8, 1635-1638

[147] Heller, J., Tudzynski, P., *Annu. Rev. Phytopathol.,* **2011**, *49*, 369-390

[148] Feng, H., Wang, X., Zhang, Q., Fu, Y., Feng, C., Wang, B., Huang, L., Kang, Z., *Biochim. Biophys. Acta,* **2014**, *1839*, 1, 1-12

[149] Kumar, A., Bimolata, W., Kannan, M., Kirti, P. B., Qureshi, I. A., Ghazi, I. A., *Funct. Integr. Genomics,* **2015**, *15*, 4, 425-437

[150] Subramanian, B., Bansal, V. K., Kav, N. N., *J. Agric. Food Chem.,* **2005**, *53*, 2, 313-324

[151] Botanga, C. J., Bethke, G., Chen, Z., Gallie, D. R., Fiehn, O., Glazebrook, J., *Mol. Plant. Microbe Interact.,* **2012**, *25*, 12, 1628-1638

[152] Maruyama, K., Urano, K., Yoshiwara, K., Morishita, Y., Sakurai, N., Suzuki, H., Kojima, M., Sakakibara, H., Shibata, D., Saito, K., Shinozaki, K., Yamaguchi-Shinozaki, K., *Plant Physiol.,* **2014**, *164*, 4, 1759-1771

[153] Dunn, M. F., Ramirez-Trujillo, J. A., Hernandez-Lucas, I., *Microbiology,* **2009**, *155*, Pt 10, 3166-3175

[154] Chen, H. Y., Huh, J. H., Yu, Y. C., Ho, L. H., Chen, L. Q., Tholl, D., Frommer, W. B., Guo, W. J., *Plant J.,* **2015**, *83*, 6, 1046-1058

[155] Hardham, A. R., Jones, D. A., Takemoto, D., *Curr. Opin. Plant Biol.,* **2007**, *10*, 4, 342-348

[156] Sun, L., Yang, D. L., Kong, Y., Chen, Y., Li, X. Z., Zeng, L. J., Li, Q., Wang, E. T., He, Z. H., *Mol. Plant Pathol.,* **2014**, *15*, 2, 161-173

[157] Etalo, D. W., Stulemeijer, I. J., Peter van Esse, H., de Vos, R. C., Bouwmeester, H. J., Joosten, M. H., *Plant Physiol.,* **2013**, *162*, 3, 1599-1617

[158] Trouvelot, S., Heloir, M. C., Poinssot, B., Gauthier, A., Paris, F., Guillier, C., Combier, M., Trda, L., Daire, X., Adrian, M., *Front. Plant Sci.,* **2014**, *5*, 592, 1-14

[159] Proels, R. K., Huckelhoven, R., *Mol. Plant Pathol.,* **2014**, *15*, 8, 858-864

[160] Watanabe, M., Balazadeh, S., Tohge, T., Erban, A., Giavalisco, P., Kopka, J., Mueller-Roeber, B., Fernie, A. R., Hoefgen, R., *Plant Physiol.,* **2013**, *162*, 3, 1290-1310

[161] Burgos, A., Szymanski, J., Seiwert, B., Degenkolbe, T., Hannah, M. A., Giavalisco, P., Willmitzer, L., *Plant J.,* **2011**, *66*, 4, 656-668

[162] Dörmann, P., *eLs,* **2013**, *January*, 1-7

[163] Welti, R., Shah, J., Li, W., Li, M., Chen, J., Burke, J. J., Fauconnier, M.-L., Chapman, K., Chye, M.-L., Wang, X., *Front. Biosci.,* **2007**, *12*, 2494-2506

[164] Pant, B. D., Burgos, A., Pant, P., Cuadros-Inostroza, A., Willmitzer, L., Scheible, W. R., *J. Exp. Bot.,* **2015**, *66*, 7, 1907-1918

[165] Natera, S. H. A., Hill, C. B., Rupasinghe, T. W. T., Roessner, U., *Funct. Plant Biol.,* **2016**, *43*, 2, 207

[166] Tenenboim, H., Burgos, A., Willmitzer, L., Brotman, Y., *Biochimie,* **2016**

[167] Kaup, M. T., Froese, C. D., Thompson, J. E., *Plant Physiol.,* **2002**, *129*, 4, 1616-1626

[168] Lin, W., Oliver, D. J., *Plant Sci.,* **2008**, *175*, 3, 233-237

[169] Okazaki, Y., Saito, K., *Plant J.,* **2014**, *79*, 4, 584-596

[170] Magnin-Robert, M., Le Bourse, D., Markham, J., Dorey, S., Clement, C., Baillieul, F., Dhondt-Cordelier, S., *Plant Physiol.,* **2015**, *169*, 3, 2255-2274

[171] Pinosa, F., Buhot, N., Kwaaitaal, M., Fahlberg, P., Thordal-Christensen, H., Ellerstrom, M., Andersson, M. X., *Plant Physiol.,* **2013**, *163*, 2, 896-906

[172] Triebl, A., Trotzmuller, M., Eberl, A., Hanel, P., Hartler, J., Kofeler, H. C., *J. Chromatogr. A,* **2014**, *1347*, 104-110

[173] Giavalisco, P., Köhl, K., Hummel, J., Seiwert, B., Willmitzer, L., *Anal. Chem.,* **2009**, *81*, 6546–6551

[174] Coutinho, I. D., Baker, J. M., Ward, J. L., Beale, M. H., Creste, S., Cavalheiro, A. J., *J. Agric. Food Chem.,* **2016**, *64*, 21, 4198-4206

[175] Schulz, M., Schnabl, H., Manthe, B., Schweihofen, B., Casser, I., *Phytochemistry,* **1993**, *33*, 2, 291-294

[176] Jensen, J. S. R. E., Petersen, B. O., Veselinovic, T., Olafsdottir, E. S., Duus, J. Ø., Omarsdottir, S., *Carbohydr. Polym.,* **2010**, *80*, 3, 799-807

[177] Shibuya, N., Minami, E., *Physiol. Mol. Plant Pathol.,* **2001**, *59*, 5, 223-233

[178] Markham, K. R., Mitchell, K. A., *Z. Naturforsch.,* **2003**, *85*, c, 53-56

[179] Weidenbörner, M., Jha, H. C., *Acta Hortic.,* **1994**, *381*, 702–708

[180] Mierziak, J., Kostyn, K., Kulma, A., *Molecules,* **2014**, *19*, 10, 16240-16265

[181] Du, Y., Chu, H., Wang, M., Chu, I. K., Lo, C., *J. Exp. Bot.,* **2010**, *61*, 4, 983-994

[182] Valares Masa, C., Sosa Diaz, T., Alias Gallego, J. C., Chaves Lobon, N., *Molecules,* **2016**, *21*, 3

[183] Peñuelas, J., Estiarte, M., Kimball, B. A., *Photosynthetica,* **1999**, *37*, 4, 615-619

[184] Yang, Z., Nakabayashi, R., Okazaki, Y., Mori, T., Takamatsu, S., Kitanaka, S., Kikuchi, J., Saito, K., *Metabolomics,* **2014**, *10*, 4, 543-555

[185] Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., Yu, S., Zhang, Q., Luo, J., *Proc Nat Acad Sci,* **2013**, *110*, 50, 20320–20325

[186] Pan, X., Welti, R., Wang, X., *Nat. Protoc.,* **2010**, *5*, 6, 986-992

[187] Okamoto, M., Tsuboi, Y., Chikayama, E., Kikuchi, J., Hirayama, T., *Plant Biotechnol.,* **2009**, *26*, 551–560

[188] van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., van der Werf, M. J., *BMC Genomics,* **2006**, *7*, 142

[189] Ernst, M., Silva, D. B., Silva, R. R., Vencio, R. Z., Lopes, N. P., *Nat. Prod. Rep.,* **2014**, *31*, 6, 784-806

[190] Tenenboim, H., Brotman, Y., *Trends Plant Sci.,* **2016**

[191] Kim, H. K., Verpoorte, R., *Phytochem. Anal,* **2010**, *21*, 1, 4-13

[192] Gibon, Y., Usadel, B., Blaesing, O. E., Kamlage, B., Hoehne, M., Trethewey, R., Stitt, M., *Genome Biol.,* **2006**, *7*, 8, R76

[193] Ayliffe, M., Periyannan, S. K., Feechan, A., Dry, I., Schumann, U., Wang, M. B., Pryor, A., Lagudah, E., *Mol. Plant. Microbe Interact.,* **2013**, *26*, 6, 658-667

[194] Mille-Lindblom, C., von Wachenfeldt, E., Tranvik, L. J., *J. Microbiol. Methods,* **2004**, *59*, 2, 253-262

[195] McGrann, G. R., Stavrinides, A., Russell, J., Corbitt, M. M., Booth, A., Chartrain, L., Thomas, W. T., Brown, J. K., *J. Exp. Bot.,* **2014**, *65*, 4, 1025-1037

[196] Spielmeyer, W., Mago, R., Wellings, C., Ayliffe, M., *BMC Plant Biol.,* **2013**, *13*, 96

[197] Joosten, M. H., *Methods Mol. Biol.,* **2012**, *835*, 603-610

[198] Shetty, N. P., Jensen, J. D., Knudsen, A., Finnie, C., Geshi, N., Blennow, A., Collinge, D. B., Jorgensen, H. J., *J. Exp. Bot.,* **2009**, *60*, 15, 4287-4300

[199] Fuhrs, H., Behrens, C., Gallien, S., Heintz, D., Van Dorsselaer, A., Braun, H. P., Horst, W. J., *Ann. Bot.,* **2010**, *105*, 7, 1129-1140

[200] Nouchi, I., Hayashi, K., Hiradate, S., Ishikawa, S., Fukuoka, M., Chen, C. P., Kobayashi, K., *Plant Cell Physiol.,* **2012**, *53*, 9, 1659-1668

[201] Krueger, S., Giavalisco, P., Krall, L., Steinhauser, M. C., Bussis, D., Usadel, B., Flugge, U. I., Fernie, A. R., Willmitzer, L., Steinhauser, D., *PLoS One,* **2011**, *6*, 3, e17806

[202] Vinaixa, M., Samino, S., Saez, I., Duran, J., Guinovart, J. J., Yanes, O., *Metabolites,* **2012**, *2*, 4, 775-795

[203] Himmelbach, A., Liu, L., Zierold, U., Altschmied, L., Maucher, H., Beier, F., Muller, D., Hensel, G., Heise, A., Schutzendubel, A., Kumlehn, J., Schweizer, P., *Plant Cell,* **2010**, *22*, 3, 937-952

[204] Hummel, J., Segu, S., Li, Y., Irgang, S., Jueppner, J., Giavalisco, P., *Front. Plant Sci.,* **2011**, *2*, 54, 1-17

[205] Csukasi, F., Osorio, S., Gutierrez, J. R., Kitamura, J., Giavalisco, P., Nakajima, M., Fernie, A. R., Rathjen, J. P., Botella, M. A., Valpuesta, V., Medina-Escobar, N., *New Phytol.,* **2011**, *191*, 2, 376-390

[206] Cuadros-Inostroza, A., *gdps* R package version 0.1.4: GeneData Processing Script, **2013**

[207] Ihaka, R., Gentleman, R., *J. Comp. Graph. Stat,* **1996**, *5*, 3, 299-314

[208] R Development Core Team, http://www.r-project.org, accessed 14.01.2014

[209] Stacklies, W., Redestig, H., Scholz, M., Walther, D., Selbig, J., *Bioinformatics,* **2007**, *23*, 9, 1164-1167

[210] Warnes, G. R., *gplots* R package version 2.11.0.1: Various R programming tools for plotting data., **2013**

[211] Hummel, J., Strehmel, N., Selbig, J., Walther, D., Kopka, J., *Metabolomics,* **2010**, *6*, 2, 322-333

[212] Benjamini, Y., Hochberg, Y., *J. R. Stat. Soc. Series B Stat. Methodol.,* **1995**, *57*, 1, 289-300

[213] Oliveros, J.C., http://bioinfogp.cnb.csic.es/tools/venny, accessed 08.02.2016

[214] Universiteit Gent, http://bioinformatics.psb.ugent.be/webtools/Venn/, accessed 02.03.2016

— Chapter 3 —

Targeted analysis of flavonols in *Arabidopsis thaliana*

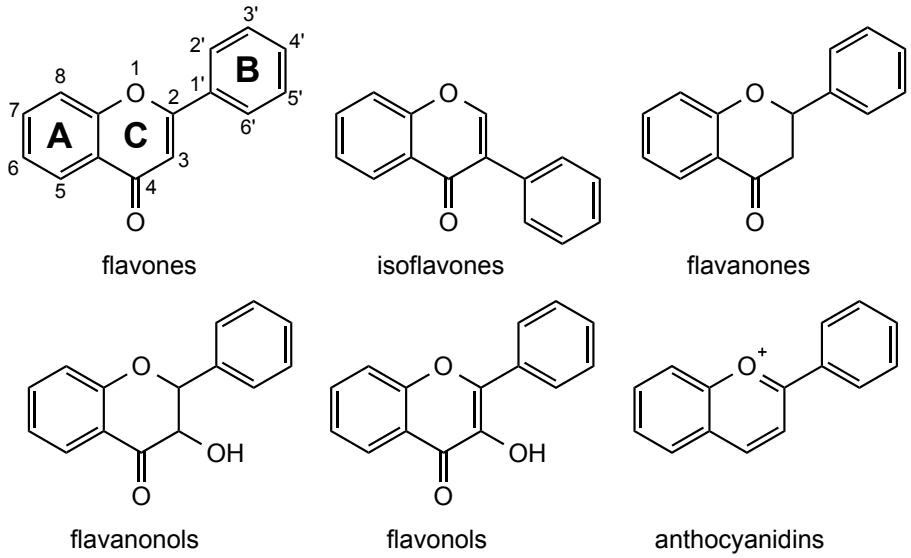
Targeted analysis of flavonols in *Arabidopsis thaliana*

For the current chapter, the numbering of chemical compounds was done independently from other chapters of this thesis.

8 Introduction

Flavonoids are a large group of polyphenolic plant secondary metabolites, comprising over 7000 different compounds [1]. The chemical core structures of these natural products are based on a three-ring diphenylpropane (C6-C3-C6) skeleton [2]. Flavonoids are classified based on these core structures, which show variations in oxidation state and degree of unsaturation of the heterocyclic ring (Scheme 13) [3]. Further structural multiplicity of flavonoids emerges from glycosylations, which represent the dominating modification, but also from backbone substituents such as hydroxyl, methoxyl or alkyl groups [2]. In principle, all hydroxyl groups can be glycosylated, but some positions are enzymatically favored, depending on the class of flavonoids. Of the enormous array of flavonoid structures, many are positional isomers.

The functions of flavonoids in plants are manifold and involve pigmentation, protection against abiotic stress factors such as UV radiation, communication with biotic factors (pathogen defense, plant-microbe communication, insect attraction) or regulation of reactive oxygen species [4, 5]. Furthermore, flavonols negatively regulate auxin transport [4, 6, 7], inhibit the cell cycle and control cell growth [4].



Scheme 13: Core scaffolds of main flavonoid classes, indicated nomenclature of rings and numbering.

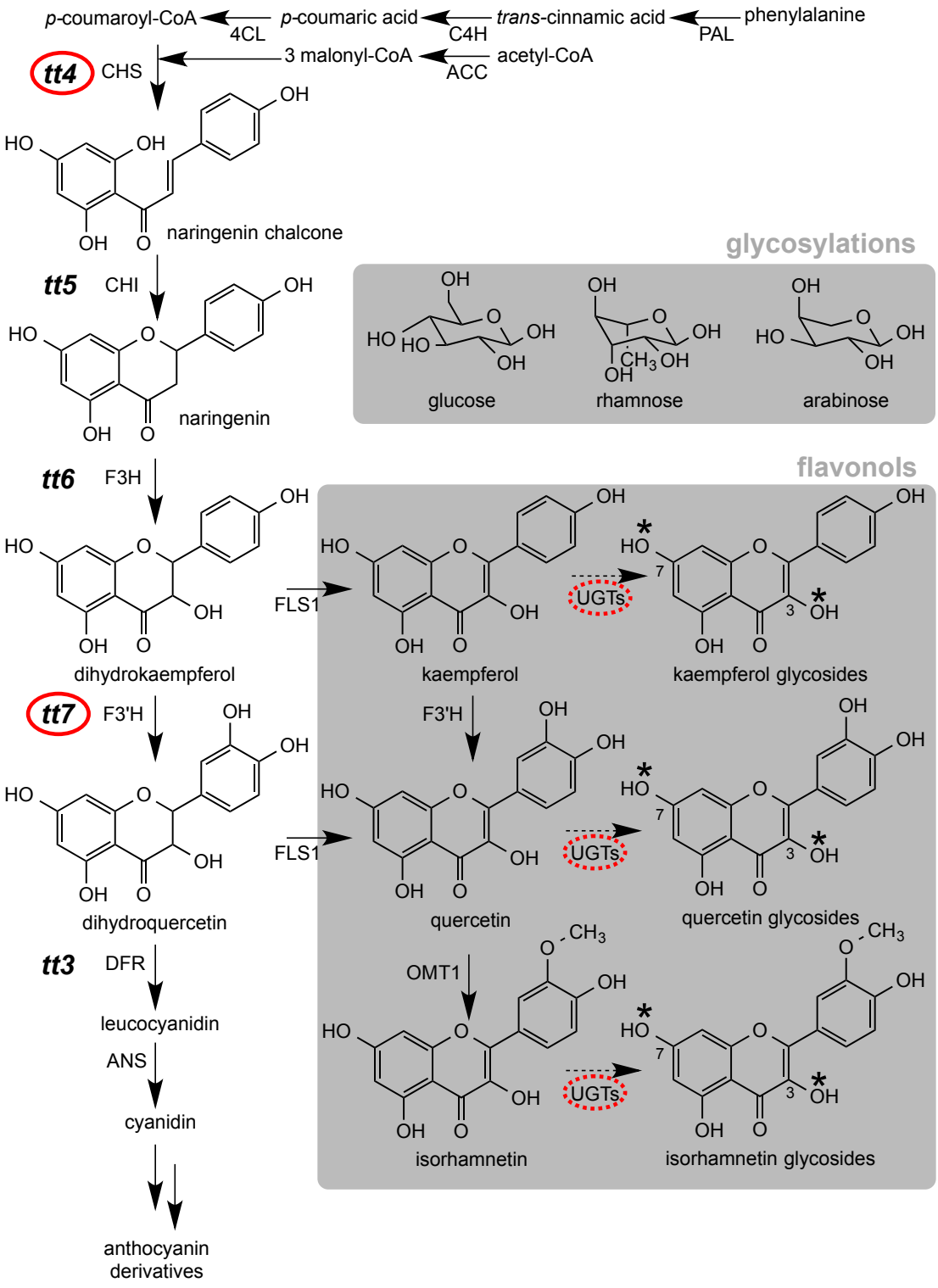
8.1 Flavonols in *Arabidopsis thaliana*

Arabidopsis thaliana (subsequently refered to as Arabidopsis) is a model plant that has been central to many studies [8], and it is the premier model plant for post genomic biology. Considered a weed, prevalent advantages of Arabidopsis are its short generation time, small plant size and efficient prolific reproduction through self-pollination. Efficient transformation procedures were established, leading to a large collection of mutants [9]. With the first plant genome to be sequenced and more than 30'000 identified genes, Arabidopsis serves as a reference for plant molecular biology, biochemistry, physiology, and development [10]. The Columbia accession is generally viewed as wild type reference genotype, and extensive physiological, biochemical, genetic and transcriptional information are available [9] which are collected in *The Arabidopsis Information Resource*, a central database for Arabidopsis research [10].

The known flavonoid content of Arabidopsis comprises mainly flavonols, but also anthocyanins and proanthocyanidins [11]. The biosynthesis of flavonols is well characterized in Arabidopsis [11]. Flavonols are synthesized via the phenylpropanoid pathway that starts with the condensation of phenylalanine derived hydroxycinnamic acids and malonyl-CoA originating from the shikimate and arogenate pathways (Scheme 14)[12]. The enzyme flavonol synthase (FLS1) produces the flavonols kaempferol and quercetin out of the corresponding dihydroflavonols, and isorhamnetin is formed by methylation of quercetin by *O*-methyltransferase (OMT1). The flavonol core scaffolds are glycosylated by UDP-dependent glycosyltransferases (UGTs) at OH-groups at positions C3 and C7. The sugar moieties attached to Arabidopsis flavonols are mainly glucose and rhamnose, rarely arabinose [11]. In *A. thaliana,* 35 flavonols were recently reported in different tissues [11] with mainly mono-, di- and triglycosylations.

A number of Arabidopsis mutants with defects in enzymes of specific flavonoid biosynthesis steps have been identified [13]. Several of the mutants phenotypically display pale-yellow seed coats, caused by the lack of proanthocyanidins, and were subsequently named *transparent testa* (*tt*) mutants [14] (Scheme 14). The *Arabidopsis thaliana transparent testa* mutant *tt4* [6] is deficient in chalcone synthase (CHS), the entry point into the flavonoid pathway, and *tt4* plants are therefore completely devoid of flavonoids [15]. In the *Arabidopsis thaliana tt7* mutant the enzyme flavonol 3'-hydroxylase (F3'H) is affected, with the consequence that *tt7* plants over-accumulate flavonols with kaempferol core but

are unable to produce quercetin and anthocyanins further downstream the biosynthesis pathway.



Scheme 14: Flavonol biosynthesis in *Arabidopsis thaliana* (adapted from [11, 16] and [17]); *transparent testa* mutants are indicated in bold and preferred glycosylation positions are marked with asterisks. Important enzymes are described in the text. Mutants used for structure elucidation are cycled in red (*tt4*, *tt7*, and *rol1-2 ugt89c1* within the UGT class).

The *Arabidopsis thaliana rol1-2* (*repressor of lrx1*) mutant [18] is affected in the rhamnose synthase (RHM1), an enzyme that converts UDP-glucose to UDP-rhamnose. The *rol1-2* mutation displays phenotypic alterations in roots (shorter root hairs, shorter roots [18]), and in shoots (hyponastic cotyledons [18], distorted adaxial cotyledon pavement cells and defective trichomes [19]). The *rol1-2* mutation had been shown to cause modification in the flavonol glycosylation profile with less rhamnosylated and more glycosylated flavonols in *rol1-2* compared to Columbia wild type [19].

The root phenotype is thought to be induced by the changes in the cell wall component pectin, which is rich in rhamnose. The shoot phenotype has been shown to be induced by the altered flavonol content as it is fully suppressed by blocking flavonol biosynthesis [17, 19]. Due to its pronounced shoot phenotype, the *rol1-2* mutant serves as a model system to investigate the mode of action of flavonols on plant development.

The glycosylation of flavonols is mainly catalyzed by several UDP-dependent glycosyltransferases (UGTs) [20]. The *rol1-2 ugt89c1* mutant is affected in the enzyme 7-*O*-rhamnosyltransferase [21] and is unable to rhamnosylate the OH-group at position C7 of flavonols.

The mutants *tt4* (defective in chalcone synthase), *tt7* (defective in flavanone-3'-hydroxylase) and *rol1-2 ugt89c1* (defective in the UDP-rhamnose: flavonol 7-*O*-rhamnosyltransferase (UGT89C1)) were used as biological tools for flavonol structure assignment.

8.2 Structural analysis of flavonoids by LC-MS

NMR spectroscopy generally is the gold standard for structural characterization of natural products, and full structural characterization of flavonoids had been performed for mono-, di- and tri-glycosides of kaempferol in *A. thaliana* [22], [23]. Due to higher sensitivity, complex plant extracts containing a variety of flavonoids at low concentrations are presently predominantly analyzed and characterized by liquid chromatography hyphenated with mass spectrometry [24].

8.2.1 Chromatography for separation of flavonoids in complex mixtures

The prevalent separation technique for flavonoids is reversed-phase (RP) chromatography based on C18 octadecyl-silica columns, combined with mobile phases typically containing acidified aqueous / organic solvent. Besides RP-LC, an emerging number of hydrophilic interaction chromatography (HILIC) methods are applied to flavonoid separations [3].

8.2.2 UV spectra to assist flavonoid characterization

Due to their absorption of UV-B light at 280–320 nm, flavonoids are effective UV filters and act as plant tissue protectors [20]. The highly conjugated core structures of flavonoids produce characteristic UV-VIS absorption spectra, which are currently mainly used to assign flavonoid subclasses. In contrast, mass spectrometry provides more structural information and is more sensitive [25]. Two maxima for the A- and B-ring at 300–380 nm and 240–285 nm are observed for flavonols. Glycosylations have little impact on UV-VIS spectra and shifted the B-ring absorbance to lower wavelength by 15–20 nm [26]. Hydroxylations generally cause a shift towards longer wavelengths. Methoxylated flavonoids show the same but less pronounced on UV spectra [3]. Sufficient chromatographic separation is required for UV-based quantitation, which is hampered by frequent co-elution of isomeric or isobaric flavonoids with similar physico-chemical properties in RP-LC [27].

8.2.3 Mass spectrometry for flavonoid characterization

Structural information can be obtained by collision-induced dissociation (CID) MS/MS or MS^n experiments in both (+) and (–)-ESI ionization mode. Negative ionization was reported to have better sensitivity, due to lower chemical noise as compared to positive ionization mode [28, 29]. (–)-ESI-MS/MS fragmentation results primarily in loss of sugar rings, but product ion spectra may not provide sufficient information to differentiate isomers [30]. Positive ionization spectra may produce more distinctive fragmentation patterns.

By the help of MS/MS spectra, extensive structural information about flavonoids can be obtained, regarding: (1) the structure of the aglycone, (2) the glycosylation types, (3) the positions of the glycosylations and (4) the nature of the glycoconjugates [2].

Additionally to flavonoid mass spectra reported in literature, a growing number of MS and MS^2 spectra can be found in electronic mass spectral depositories such as *ReSpec*t [31], *MassBank* [32] or *Phenol-Explorer* [33].

8.2.4 Structure of flavonoid aglycone

Generally, the aglycone cores undergo reproducible fragmentations and can be assigned by comparing the spectra of glycosylated flavonoids with spectra from native aglycones. In principle, the fragmentation rules are applicable to both (+)- and (–)-ESI mode, although higher collision energies are required in negative mode, and diagnostic ions might be missing [34, 35].

8.2.5 Flavonoid glycosylation type

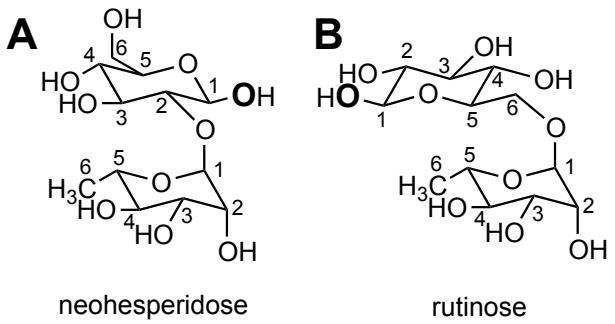
In glycosylated flavonols, the *O*-glycosidic bond is already cleaved under low collision energy conditions. As a consequence, the (–)-ESI CID spectra of flavonoid-*O*-glycosides reveal the attached saccharide units as neutral losses of consecutive fragments, calculated as the mass difference between the *m/z* values of the ionized molecule and the aglycone ions [26, 34]. For example, hexoses display a neutral loss of 162 Da, but isomers like glucose and galactose cannot be distinguished. The neutral loss of 146 Da is indicative for rhamnose, as it is the only deoxyhexose known in flavonoid conjugates [28].
Besides *O*-glycosylated flavonoids, several plant species also accumulate *C*-glycosides (e.g. maize, wheat and rice [36]), in which the sugar unit is attached through a less reactive carbon-carbon bond. Rupture of the sugar ring results in characteristic mass spectra with neutral losses of 120 and 122 Da in case of hexose and deoxyhexose units [2], which are distinguishable from fragments of *O*-glycosides. Arabidopsis was long thought to lack the biosynthetic pathway for flavones, the main *C*-glycoside containing flavonoid class [37], but recently a respective enzyme had been reported [38], and the flavone apigenin was detected in cauline and senescing leaves.

8.2.6 Flavonoid glycosylation position

For flavonol core structures, mainly the OH groups in positions C-3 and C-7 are glycosylated (Scheme 1). The bond strength, and subsequently the prevalence for cleavage, depends on the glycosylation position. Glycans attached to 3-*O*-position were reported to be cleaved preferentially in spectra of $[M+H]^+$ [39].
For monoglycosides, the $[aglycone–CO–H]^-$ ions are characteristic for 7-*O*-glycosylations, whereas $[aglycone–2H–CO–H]^-$ ions are specific for 3-*O*-glycosides [40].
For flavonol-3,7-di-*O*-glycosides with different saccharide residues, glycosylation positions can be assigned by the radical ion $[aglycone+glycoside–2H]^{\bullet-}$, which originates from homolytic cleavage of the glycosidic bond [34] and only occurs in the case of 3-*O*-glycosylation [2, 41].

8.2.7 Nature of flavonoid glyco-conjugates

Diglycosides occur in different isoforms: as di-*O*-glycoside, mono-*O*-glycosylglycoside, or *C*,*O*-diglycosides [42]. Although *C*-glycosides are not observed within the flavonol subclass, complexity in flavonol structure is observed regarding interglycosidic linkage and glycan sequence.



Scheme 15: Isomeric diglycosides: neohesperidose [rhamnosyl(α1→2)glucose] (A) and rutinose [rhamnosyl(α1→6)glucose] (B), flavonol-conjugating *O*-atoms are indicated in bold.

In CID MS/MS spectra of sodiated molecules $[M+Na]^+$, the sodium ions tend to bind to the saccharide part, providing structural information about the size of the sugar moieties attached to the aglycone [27, 39].

Discrimination of mono-*O*-diglycosides and di-*O*-glycosides based on the relative intensities of fragments of CID MS/MS spectra in (+)- and (–)-ESI is difficult, because the instrumental settings of the mass spectrometer like collision energy strongly influence the fragmentation of flavonoids [43].
However, several studies were reported in the literature to either characterize glycan sequence or interglycosidic linkage, for example using two common flavonoid diglycosides containing rhamnose and glucose: neohesperidose with interglycosidic 1→2 linkage and rutinose with 1→6 linkage (Scheme 15). Interestingly, these diglycosides undergo in (+)-ESI CID at low collision energy of 10 eV a rearrangement reaction, in which the inner glucose residue is lost [34, 44]. A classification scheme to differentiate between neohesperiosides and rutionsides was suggested [45]. In (+)-ESI, the ratio of $[aglycone+H]^+$ / $[aglycone+glycoside+H]^+$ was reported to be consistently higher for 1→2 linked isomers than for 1→6 linkage [44, 46]. Also in (–)-ESI, structural studies have been performed to distinguish between different interglycosidic linkage [40], [47], [48].

8.2.8 Further techniques for flavonoid characterization

As an extension of conventional liquid chromatography, multidimensional LC combines multiple orthogonal one-dimensional separation methods to achieve higher separation power and selectivity [3]. For example, on-line two-dimensional LC coupling of HILIC with RP-LC was applied to procyanidin flavonoid analysis in grape seeds [49, 50] and to different flavonol classes in apples [51].

A complementary approach for the characterization of isobaric flavonoid isomers is ion mobility spectrometry (IMS), where ions are separated within milliseconds according their size/charge ratio and their cross sections [52]. IMS measures the drift time for an individual ion to cross a chamber filled with a buffer gas in the presence of an electric field [3]. IMS-TOF-MS successfully separated isobaric flavonoid standards [53]. In combination with chromatographic separation, flavonoids were characterized with IMS in complex mixtures such as cauliflower waste, [54], in black tea [55] or in urine [56].

9 Aim

In the past, a HPLC-MS method for the targeted analysis of flavonol profiles in Arabidopsis extracts has been established in our laboratory [19]. To profit from advanced technology, the method should be transferred from a conventional HPLC – ion trap MS instrument [19] to a ultra-high performance liquid chromatography (UHPLC) – high resolution QTOF mass spectrometer [17]). After method transfer, a reassignment of the flavonol structures should be performed. The optimized method is expected to be applied to various biological questions regarding the function of flavonols.

10 Flavonol structure assignment

The method transfer involved changes on both the LC- and MS-part of the method.
On one hand, the previous method based on conventional HPLC was advanced to state-of-the-art ultra-high performance liquid chromatography. By reducing the particle size of the column packing material from 3 µm (HPLC) to < 2 µm (UHPLC) benefits in sensitivity, selectivity and speed are achieved. Based on terms of the van Deemter curve, lower Eddy diffusion and resistance to mass transfer combined with increased linear velocity of the mobile phase culminate in a lower minimum plate height and a higher separation efficiency for a given column length. Increased separation speed is achieved by a combination of shorter column length and higher flow rate. Increased flow rates limit the longitudinal diffusion, sharpen the peak shape and thus increase sensitivity. The advanced resolution and higher peak capacity ameliorate separation selectivity. A disadvantage of small particles is the concomitant increase of the backpressure to above 400 bars, which is manageable by dedicated high-pressure instruments [3]. For our investigation, an UHPLC column (*Acquity BEH C18, Waters*) was selected that was packed with a stationary phase possessing a very similar polarity as the previously used HPLC column (*Nucleosil C18, Macherey–Nagel*), also fully endcapped, and to be especially robust (trifunctional ligand

binding) according to current commercial technology. To further profit from the advantages arising from UHPLC, a shorter gradient with steeper slope was designed.
On the other hand, the MS-method was transferred from an ion-trap mass spectrometer to a QTOF instrument. A major benefit of the method transfer was the accurate mass information obtained from the QTOF instrument, whereas the previously used ion trap mass spectrometer was only capable of nominal mass measurements. Accurate masses allow the distinction of compounds with chemical formulas of similar masses, based on the accurate measurement of the *m/z* ratio up to five digits. For example, the Arabidopsis metabolites neoglucobrassicin $C_{17}H_{22}N_2O_{10}S_2$ with $[M–H]^-$ *m/z* 477.06431 and isorhamnetin-3-*O*-glucoside $C_{22}H_{22}O_{12}$ with $[M–H]^-$ *m/z* 477.10385 could well be distinguished in extracted ion chromatograms (EICs) with an isolation width of 0.01 Da.
Following the method transfer to the shorter UHPLC method, the flavonol structures were reassigned by applying several parameters like retention time, UV spectra, peak abundance in extracts of different biological mutants affected in their flavonoid biosynthesis, mass accuracy and CID MS/MS spectra.

The general naming convention for flavonoids is based on a combination of the aglycone, the type of linkage atom (*C*, *O*), the glycosylation position, and the name of the saccharide(s). For flavonols described in this chapter, only *O*-glycosylations were detected, and therefore the type of linkage was omitted (e.g. K-G-3 for kaempferol-3-*O*-glycoside). Three types of assignments of flavonol glycosides were obtained, in the following exemplified for diglycosides. The first type includes diglycoside flavonols where the glycoside sequence could be determined with a certain degree of confidence, and the sugar unit directly attached to the flavonol core was emphasized by underlining (e.g. Q-R<u>G</u>-3). The second type included assingements where a diglycoside was detected but the glycoside sequence remained unknown, (e.g. Q-GR) and the third type involved flavonol assignments for which the glycosylation positions could not be determined (e.g. Q+G+R).

10.1 Flavonol retention times

For the transferred method, the elution order of the flavonols was fully conserved between the earlier HPLC and the shorter UHPLC method (Figure 1 and Table 1). In RP chromatography, polar compounds elute at the beginning of the chromatogram. Within the same phenol class, decreasing retention times were reported for compounds with more hydroxyl groups [26]; this tendency in retention time was confirmed in the sense that

quercetin flavonols eluted earlier than the analogous flavonols with kaempferol aglycones (e.g. Q-G-3 at 5.29 min (flavonol **23**) versus K-G-3 at 5.96 min (flavonol **27**)).

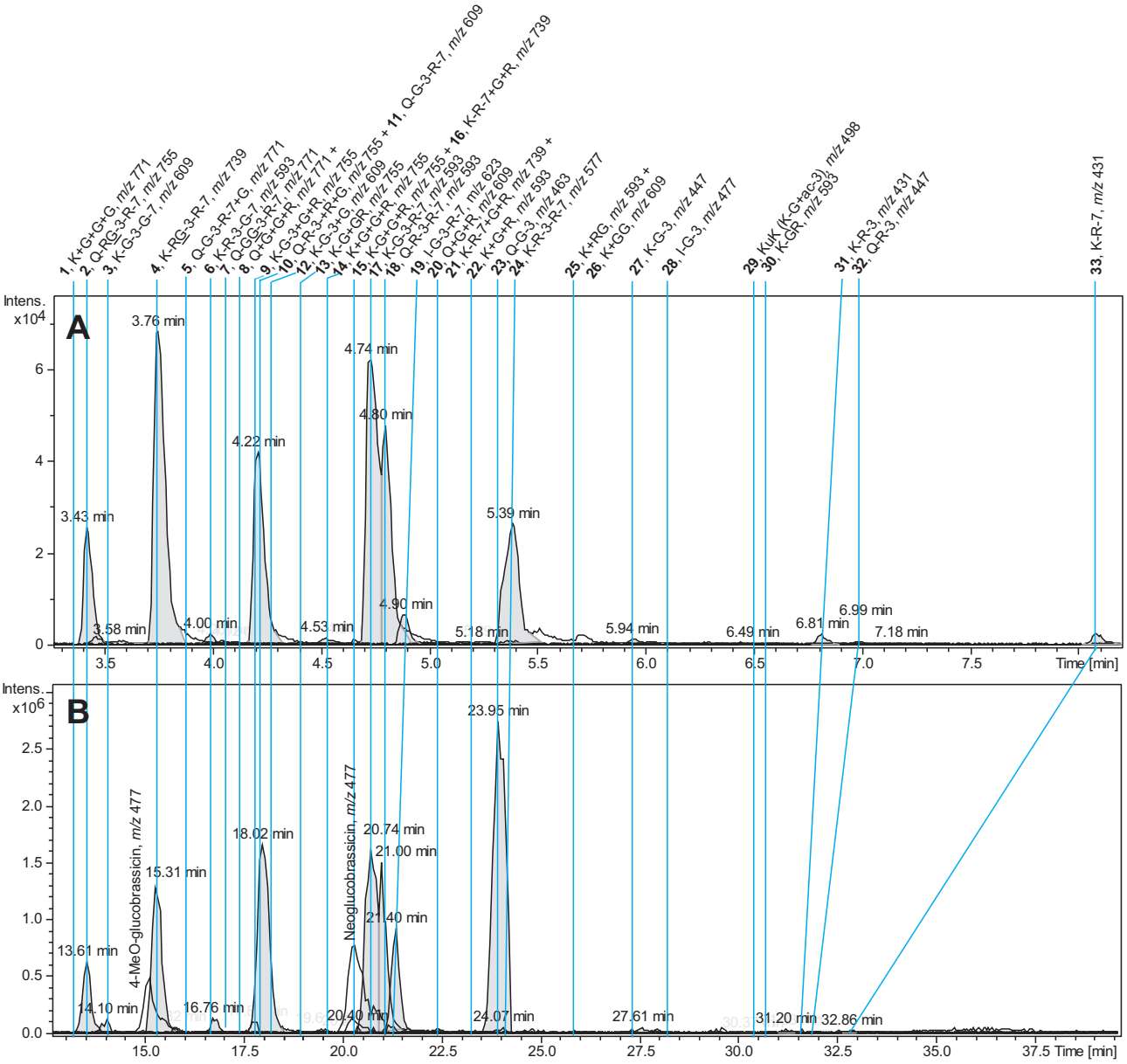


Figure 1: Correlation of Arabidopsis Columbia flavonol profiles on different instruments: UHPLC (A) and the older HPLC method (B).

Glycosylated polyphenols had been described in literature to elute in advance to the corresponding aglycones [26]. With the current method, tri-glycosylated flavonols eluted in advance of their di- and mono-glycosylated analogs, but aglycones were not detected. Glycosylated flavonoids were reported to elute in RP-LC generally in the following order: galactosides < glucosides < arabinosides < xylosides< rhamnosides [57]. The tendency of later elution of rhamnoslyated flavonols could be confirmed (e.g. K-G-3 at 5.96 min (flavonol **27**) versus K-R-3 and K-R-7 at 6.83 and 8.09 min respectively (flavonol **31** and **33**)).

Table 1: Putative flavonols with retention times of UHPLC and HPLC methods, chemical formula, mass accuracy and peak areas of the EICs of the corresponding $[M–H]^-$ in wild type Columbia, presence in Arabidopsis *tt7* mutant, aglycones, glycosylation, Arabidopsis *rol1-2 ugt89c1* mutant, maximum UV absorptions (uncertain values indicated in grey), and references from literature (a [11], b [58], c [27], d [22], e [59], f [60], g [39], h [61], i [62], j [20]). Table sorted according to decreasing retention time.

No.	Putative flavonol	MW [g/mol]	RT UHPLC [min]	RT HPLC [min]	Chemical formula	*m/z* $[M-H]^-$ calculated	*m/z* $[M-H]^-$ measured	Mass accuracy [ppm]	Peak area Columbia root	Peak area Columbia shoot	*tt7*	Aglycone	Glyco-syl units	*rol1-2 ugt89c1*	UV_{max} [nm]	Reference *Arabidopsis*
1	K+G+G+G	772	3.35	13.09	$C_{33}H_{40}O_{21}$	771.19893	771.20044	-1.96	668	386	+	K	G, G, G	+	250, 286, 327	
2	Q-RG-3-R-7	756	3.44	13.61	$C_{33}H_{40}O_{20}$	755.20401	755.20310	1.22	57667	51282	–	Q	G, R, R	–	252, 347	f
3	K-G-3-G-7	610	3.48	14.10	$C_{27}H_{30}O_{16}$	609.14610	609.14590	0.34	3076	8239	+	K	G, G	+	257, 347	a, b, d, c
4	K-RG-3-R-7	740	3.78	15.31	$C_{33}H_{40}O_{19}$	739.20910	739.20867	0.58	105883	174591	+	K	G, R, R	–	265, 340	g, h
5	Q-G-3-R-7+G	772	3.85	15.82	$C_{33}H_{40}O_{21}$	771.19893	771.19722	2.22	9047	628	–	Q	G, G, R	–	265, 285,330	
6	K-R-3-G-7	594	3.98	16.76	$C_{27}H_{30}O_{15}$	593.15119	593.15058	1.04	2255	5072	+	K	G, R	+	264, 327	a, b, c
7	Q-GG-3-R-7	772	4.05	17.40	$C_{33}H_{40}O_{21}$	771.19893	771.19788	1.36	1220	775	–	Q	G, G, R	–	263,289, 326	g, i, j
8	Q+G+G+R	772	4.12	17.80	$C_{33}H_{40}O_{21}$	771.19893	771.19887	0.08	809	2069	–	Q	G, G, R	+	n/a	
9	K-G-3+G+R	756	4.12	n/a	$C_{33}H_{40}O_{20}$	755.20401	755.20244	2.08	212	1382	+	K	G, G, R	+	n/a	
10	Q-R-3+R+G	756	4.20	n/a	$C_{33}H_{40}O_{20}$	755.20401	755.20497	-1.27	13893	336	–	Q	G, R, R	+	270, 349	c
11	Q-G-3-R-7	610	4.20	18.02	$C_{27}H_{30}O_{16}$	609.14610	609.14610	0.02	330925	127387	–	Q	G, R	–	255, 352	g, h
12	K-G-3+G	610	4.25	n/a	$C_{27}H_{30}O_{16}$	609.14610	609.14514	1.58	*	*	+	K	G, G	+	255, 353	
13	K+G+GR	756	4.30	n/a	$C_{33}H_{40}O_{20}$	755.20401	755.20273	1.69	1254	359	+	K	G, G, R	–	245, 288, 337	c
14	K+G+G+R	756	4.53	19.61	$C_{33}H_{40}O_{20}$	755.20401	755.20191	2.78	1615	3602	+	K	G, G, R	–	272, 289, 335	c
15	K+G+G+R	756	4.68	n/a	$C_{33}H_{40}O_{20}$	755.20401	755.20235	2.20	444	1382	+	K	G, G, R	–	n/a	
16	K-R-7+G+R	740	4.70	20.36	$C_{33}H_{40}O_{19}$	739.20910	739.20690	2.98	15051	1629	+	K	G, R, R	–	266, 336	

* Only detected in *tt7* / *rol1-2 ugt89c1* mutants

Table 1, *continued*

No.	Putative flavonol	MW [g/mol]	RT UHPLC [min]	RT HPLC [min]	Chemical formula	*m/z* [M-H]$^-$ calculated	*m/z* [M-H]$^-$ measured	Mass accuracy [ppm]	Peak area Columbia root	Peak area Columbia shoot	*tt7*	Aglycone	Glyco-syl units	*rol1-2 ugt89c1*	UV_{max} [nm]	Reference *Arabidopsis*
17	K-G-3-R-7	594	4.74	20.74	$C_{27}H_{30}O_{15}$	593.15119	593.15136	-0.27	255001	195264	+	K	G, R	–	265, 346	a, b, g, h
18	Q-R-3-R-7	594	4.81	21.00	$C_{27}H_{30}O_{15}$	593.15119	593.15015	1.77	50031	95623	–	Q	R, R	–	265, 315	g, h
19	I-G-3-R-7	624	4.90	21.40	$C_{28}H_{32}O_{16}$	623.16175	623.16210	-0.54	99791	33450	–	I	G, R	–	254, 351	d, e, g, h
20	Q+G+R	610	5.03	22.40	$C_{27}H_{30}O_{16}$	609.14610	609.14525	1.40	8009	1829	–	Q	G, R	+	286, 325	
21	K-R-7+G+R	740	5.20	n/a	$C_{33}H_{40}O_{19}$	739.20910	739.20754	2.11	240	1459	+	K	G, R, R	–	n/a	e
22	K+G+R	594	5.20	n/a	$C_{27}H_{30}O_{15}$	593.15119	593.14995	2.09	705	1658	+	K	G, R	+	n/a	
23	Q-G-3	464	5.29	24.07	$C_{21}H_{20}O_{12}$	463.08820	463.08858	-0.83	5916	20304	–	Q	G	+	251, 264, 337	a, g
24	K-R-3-R-7	578	5.37	23.95	$C_{27}H_{30}O_{14}$	577.15628	577.15619	0.16	34935	74996	+	K	R, R	–	265, 322	a, b, g, h
25	K+RG	594	5.69	n/a	$C_{27}H_{30}O_{15}$	593.15119	593.15185	-1.11	2582	2033	+	K	G, R	+	265, 338	
26	K+GG	610	5.70	n/a	$C_{27}H_{30}O_{16}$	609.14610	609.14713	-1.69	222	83	+	K	G, G	+	264, 320	a, b, c, d
27	K-G-3	448	5.96	27.61	$C_{21}H_{20}O_{11}$	447.09328	447.09347	-0.41	5897	3272	+	K	G	+	265, 322	a, b, g
28	I-G-3	478	6.10	27.90	$C_{22}H_{22}O_{12}$	477.10385	477.10350	0.73	5112	13400	–	I	G	+	285, 329	a
29	K-G+ac-3	490	6.50	30.12	$C_{23}H_{22}O_{12}$	489.10385	489.10361	0.49	28	580	+	K	acetyl-G	+	265, 329, 349	a, b, c, d
30	K+G+R	594	6.55	n/a	$C_{27}H_{30}O_{15}$	593.15119	593.14969	2.53	**	**	+**	K	G, R	+	265, 328	
31	K-R-3	432	6.83	31.20	$C_{21}H_{20}O_{10}$	431.09837	431.09823	0.33	4020	4940	+	K	R	+	264, 325	a
32	Q-R-3	448	6.99	31.30	$C_{21}H_{20}O_{11}$	447.09328	447.09215	2.53	28645	1673	–	Q	R	+	285, 327	b, a
33	K-R-7	432	8.09	32.86	$C_{21}H_{20}O_{10}$	431.09837	431.09802	0.82	35111	7412	+	K	R	–	264, 357	a, c

** Only detected in protoplasts.

10.2 UV spectra

UV spectra are characteristic for each flavonoid class and mainly used to distinguish different types of flavonoids. Generally, flavonols show two UV absorption maxima at around 240–285 nm and around 300–380 nm (Table 1, UV_{max}). The flavonol cores kaempferol and quercetin were reported to display distinct maxima caused by the B-ring absorption at around 265 nm (kaempferol) and at around 256 nm (quercetin) [63]. For kaempferol-3-*O*-glucoside (K-G-3, flavonol **27**) a UV_{max} at 265 nm was observed, whereas quercetin-3-*O*-glucoside (Q-G-3, flavonol **23**) displayed a maximum at 251 nm. Yet, the B-ring absorbance was not unambiguously assigned for all flavonols, likely due to co-eluting compounds in the plant extracts. Information content of UV spectra was limited for co-eluting flavonols like K-G-3-R-7 and Q-R-3-R-7 that were not chromatographically resolved and no pure chromophore was measured in UV detection, and also for low-abundant flavonols with low absorption intensity, and according values are marked in grey.

10.3 Biological mutants

In addition to analytical tools, Arabidopsis mutants *tt4*, *tt7* and *rol1-2 ugt89c1* were used to verify structural characteristics of flavonols. All flavonols were confirmed to be absent in extracts of *tt4* plants, as these mutants lack the key enzyme in flavonol biosynthesis (Scheme 14). Flavonols present in extracts of the kaempferol over-accumulator mutant *tt7* were reasoned to contain kaempferol cores, absent peaks were attributed to quercetin or isorhamnetin flavonols (Table 1, *tt7*). The *rol1-2 ugt89c1* mutant affected in the enzyme 7-rhamnosyltransferase [21] is unable to rhamnosylate the OH-group at position C7 of flavonols. The absence of the according flavonols in extracts of *rol1-2 ugt89c1* compared to wild type was therefore used to distinguish 7-*O*-rhamnosylation from 7-*O*-glucosylation (Table 1, *rol1-2 ugt89c1*). Examples of flavonol profiles in extracts of Columbia wild type and mutants *rol1-2* and *rol1-2 ugt89c1* are shown in Figure 2.

10.4 Calculation of chemical formulas

Accurate masses of the HR-MS spectra provided information about the chemical formulas of the target flavonol glycosides. The mass accuracy for all flavonols was calculated based on HR-MS-data in negative ionization mode. All investigated flavonols fit the expected, calculated *m/z* value by less than 3 ppm mass deviation (Table 1, mass accuracy). Low-abundant signals caused by low-abundant components are typically less accurate.

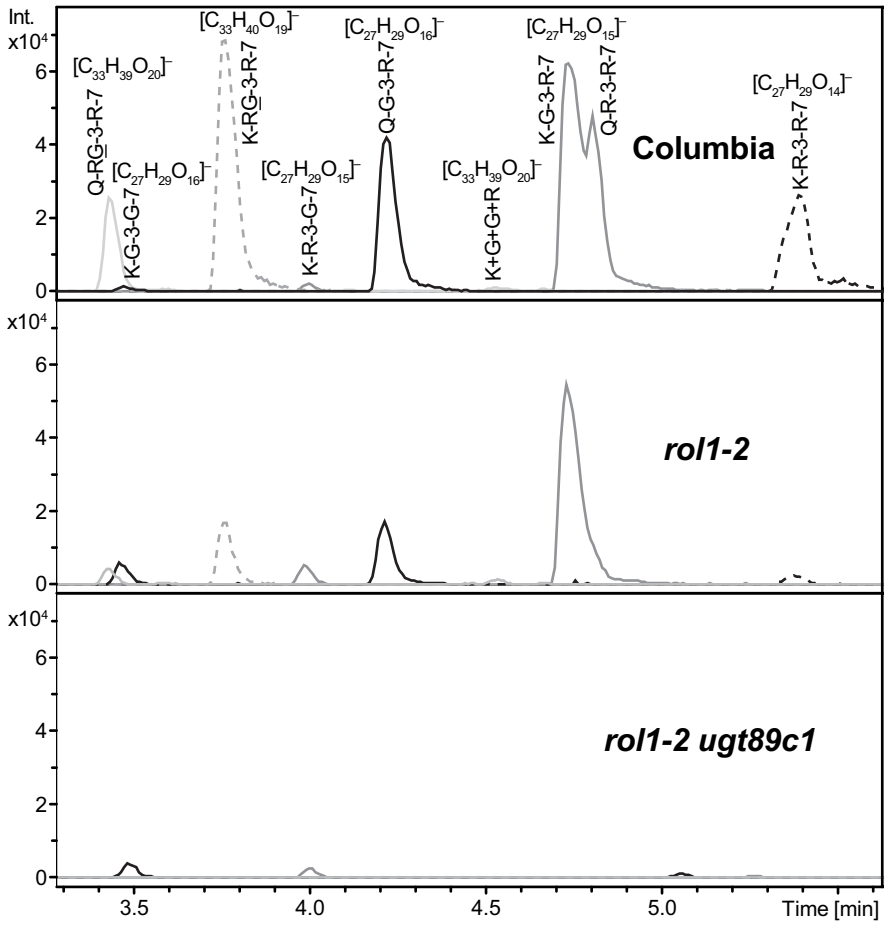


Figure 2: Flavonol profiles of Arabidopsis Columbia wild type (top) and mutants *rol1-2* (middle) and *rol1-2 ugt89c1* (bottom); different grey shades and lines indicate the individual flavonol $[M–H]^-$ EICs ± 0.02.

10.5 CID MS/MS spectra

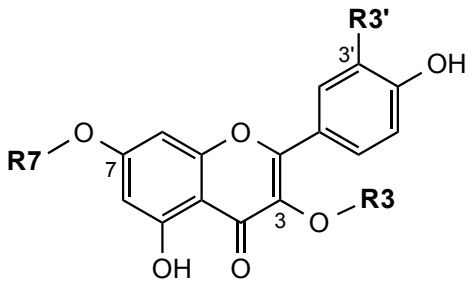
Finally, the flavonol structures were also reassigned from a comprehensive analysis of mass fragmentation patterns obtained by second-order CID MS/MS spectra in the negative and/or positive ionization mode based on $[M–H]^-$, $[M+H]^+$ and $[M+Na]^+$ ions, of which the fragments are listed in Table 2. The MS/MS spectra are shown in the supplementary information (Supplementary figure 32 and following). Different fragmentation mechanisms generate different types of ions, and thus a combination of the spectra recorded in positive and negative ionization modes may provide complementary information for improved compound identification. Post-column addition of sodium formiate solution increased the abundance of the majority of the studied sodium adducted $[M+Na]^+$ flavonol ions to a sufficient level for performing CID-MS/MS experiments, thus providing additional information about the glycoside sequences. Spectra were interpreted and compared with previously reported data from literature. In the following, the results are described in sequential order of mono-, di- and tri-glycosylated flavonols.

Table 2: MS/MS fragments of putatively annotated flavonols in *Arabidopsis thaliana*; unknown ions are colored in grey.

No.	Putative flavonol	RT $_{UHPLC}$ [min]	Main fragments $[M-H]^-$	Main fragments $[M+H]^+$	Main fragments $[M-Na]^+$	Glycoside $[x + Na]^+$
1	K+G+G+G	3.35	771, 609, 447, 285	773, 611, 449, 287	n/a	n/a
2	Q-RG-3-R-7	3.44	755, 609, 446,299	757, 611, 449, 303	779, 633, 331, 185	GR, G
3	K-G-3-G-7	3.48	609, 447, 285	611, 465, 449, 303, 287	633, 471, 309	none
4	K-RG-3-R-7	3.78	739, 593, 431, 430,284	741, 595, 433, 287	763, 617, 471, 331	GR
5	Q-G-3-R-7+G	3.85	771, 625, 609, 463, 302	773, 611, 449, 303	795, 694, 409, 185	G
6	K-R-3-G-7	3.98	593, 464, 447, 285	595, 449, 287	617, 471, 455, 309	none
7	Q-GG-3-R-7	4.05	771, 463, 447, 301, 301	773, 611, 449	795, 694, 347,185	GG, G
8	Q+G+G+R	4.12	771,447, 301	773, 611, 449, 303	795, 694, 409, 331, 185	GR, G
9	K-G-3+G+R	4.12	755, 593, 447, 369, 285	n/a	n/a	n/a
10	Q-R-3+R+G	4.20	755, 609, 463,, 446, 301	757, 611, 595, 449	779, 633, 487, 347	GG
11	Q-G-3-R-7	4.20	609, 463, 446, 301	611, 449, 303	633, 487, 325, 185	G
12	K-G-3+G	4.25	609, 446, 283	n/a	n/a	n/a
13	K+G+GR	4.30	755, 609, 582, 561, 522, 480, 471, 297, 285	n/a	n/a	n/a
14	K+G+G+R	4.53	755, 609, 447, 431, 285	757, 611, 595, 433, 287	n/a	n/a
15	K+G+G+R	4.68	755, 609, 447, 431, 285	n/a	n/a	n/a
16	K-R-7+G+R	4.70	739, 593, 431, 285	741, 595, 433, 399, 237, 160	763, 617, 471, 331	GR
17	K-G-3-R-7	4.74	593, 447, 431, 430, 283	595, 433, 287	617, 471, 309, 185	G
18	Q-R-3-R-7	4.81	593, 447, 446, 301	595, 449, 303	617, 471, 325, 169	R
19	I-G-3-R-7	4.90	623, 477, 461, 460, 445, 315, 314, 313	625, 479, 463, 317	647, 501, 485, 339, 185	G
20	Q+G+R	5.03	609, 300	611, 449, 303	633, 487, 325, 226, 185	G
21	K-R-7+G+R	5.20	739, 647, 593, 431, 285	741, 595, 579, 433, 287	763, 617, 471, 455, 331, 309	GR
22	K+G+R	5.20	593, 284	595, 433, 287	617, 471, 309, 185	G
23	Q-G-3	5.29	463, 301, 300, 271	465, 303	487, 324, 185	G
24	K-R-3-R-7	5.37	557, 431, 430, 285, 284	579, 433, 287	601, 455, 309, 169	R
25	K+RG	5.69	593, 285	595, 449, 287	617, 471, 331, 309, 185	GR, G
26	K+GG	5.70	609, 285	n/a	n/a	n/a
27	K-G-3	5.96	447, 285, 284, 255	449, 287	471, 308	n/a
28	I-G-3	6.10	477, 315, 314, 300, 299, 285	479, 317	501, 338, 185	G
29	K-G+ac-3	6.50	489, 285, 284, 255	n/a	n/a	n/a
30	K+G+R	6.55	593, 357, 389, 285	595, 561, 449, 433, 287	617, 471, 309	none
31	K-R-3	6.83	431, 285, 284, 257, 255	433, 287	n/a	n/a
32	Q-R-3	6.99	447, 301, 300, 279	449, 303	n/a	n/a
33	K-R-7	8.09	431, 285, 284, 257	433, 287	n/a	n/a

10.5.1 Monoglycosylated flavonols

The monoglycosylated flavonols **23**, **27**, **28** and **31**–**33** (Scheme 16) were assigned based on the presence or absence of the corresponding EIC peaks in the *rol1-2 ugt89c1* mutant and on the abundance of the $[K/Q–CO]^-$ fragment ions for 7-*O*-glycosylations and the $[K/Q–CH_2O]^-$ fragment ions for 3-*O*-glycosylations [2].



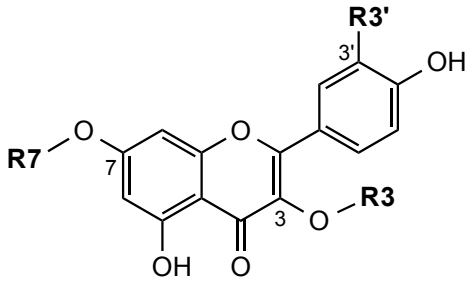
Putative flavonol	No.	R3'	R3	R7
Q-G-3	**23**	OH	G	H
K-G-3	**27**	H	G	H
I-G-3	**28**	OCH_3	G	H
K-G+ac-3	**29**	H	G+ac	H
K-R-3	**31**	H	R	H
Q-R-3	**32**	OH	R	H
K-R-7	**33**	H	H	R

Scheme 16: Monoglycosylated flavonols

Flavonol **29** with chemical formula $C_{23}H_{22}O_{12}$ was previously suggested to be kaempferol-3-*O*-(6”-acetyl-glucoside) [19], and glycosylation at 3-*O*-position was confirmed by the observed fragment $[K–CH_2O]^-$ at *m/z* 255. In the (–)-ESI-MS/MS spectrum of kaempferol-3-*O*-(6”-acetyl-glucoside), the aglycone fragment $[K–2H]^{\bullet -}$ at *m/z* 284 had been reported as the base peak in spectra from literature [64], which fits the observation for flavonol **29**. As expected, the less polar acetylated flavonol eluted slightly later from the RP-column than the kaempferol-3-*O*-glucoside analog (flavonol **29** at 6.50 min versus flavonol **27** at 5.96 min), which was also in accordance with earlier reports [63]. Previously, kaempferol-3-*O*-(6"-*O*-acetyl)-*β*-D-glucopyranoside had been found in a variety of plants, for example in male flowers of *Eucommia ulmoides* [65], in the traditional Chinese herbal medicine halophyte *Apocynum venetum* L. [66, 67], in *Morus alba* leaves extracts [63], in *Helichrysum zivojinii* [68], in strawberry and wild strawberry [69], in extracts of *Petasites japonicus* [64], in *Securigera varia* [70], in *Chorisia chodatii* flowers [71] and in needles of the norway spruce *Picea abies* [72]. Considering the broad variety of plant species in which kaempferol-3-*O*-(6”-acetyl-glucoside) had been detected, the unknown flavonol **29** in Arabidopsis might well be this compound, although the position of the acetyl group could not be confirmed based on the MS^2 data.

10.5.2 Diglycosylated flavonols

The diglycosylated flavonol **24** with chemical formula $C_{27}H_{30}O_{14}$ (Scheme 17) was appointed K-R-3-R-7 due to its absence in *rol1-2 ugt89c1* and by comparison with the reference spectrum of K-R-3-R-7 in Arabidopsis [22].



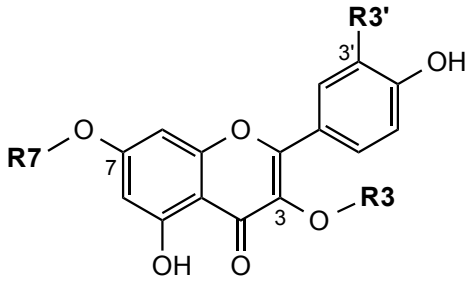
Putative flavonol	No.	R3'	R3	R7
I-G-3-R-7	**19**	OCH_3	G	R
K-R-3-R-7	**24**	H	R	R

Scheme 17: Diglycosylated flavonols **19** ($C_{28}H_{32}O_{16}$) and **24** ($C_{27}H_{30}O_{14}$)

For the isorhamnetin flavonol **19** with formula $C_{28}H_{32}O_{16}$, typical aglycone fragments at *m/z* 315 / 317 were observed in (–) and (+)-ESI-MS/MS spectra respectively. The prevalent loss of the glucoside unit pointed towards 3-*O*-glucosylation, and spectra showed high similarity to I-G-3-R-7 in references [60, 62]. An additional indication for the 3-*O*-glycosylation was the radical ion $[I+R–2H]^{\bullet-}$, which is typical for the homolytic cleavage of the 3-*O*-glycosidic bond [2]. The assignment of the 7-*O*-rhamnosylation was further supported by the lack of a corresponding signal in the rhamnosyltransferase mutant.

Of the six diglycosylated flavonoids with formula $C_{27}H_{30}O_{15}$ (Scheme 18), flavonols **6** and **17** were putatively identified as kaempferol 3,7-di-*O*-glycosides (Figure 3). The ions $[K–2H–H]^-$ observed for both flavonols **6** and **17** were previously reported as typical for 3,7-di-*O*-glycosides [40]. Flavonol **6** was present in the *rol1-2 ugt89c1* mutant, and showed prevalent fragments at *m/z* 447 / 449 in (–)- and (+)-ESI-MS/MS corresponding to the neutral loss of a rhamnoside moiety, and was accordingly assigned as K-R-3-G-7. In contrast, flavonol **17** was absent in the *rol1-2 ugt89c1* mutant, suggesting 7-*O*-rhamnosylation, and showed fragments *m/z* 431 / 433 after neutral loss of a glycosyl unit. Thus, flavonol **17** was assigned as K-G-3-R-7. The assignments were further supported by the radical ions $[K+G–2H]^{\bullet-}$ and $[K+R–2H]^{\bullet-}$ formed by the elimination of a glucosyl radical, which was exclusively observed for the homolytic cleavage of the 3-*O*-glycosidic bond [2]. Fragmentation of the sodiated molecules showed similar spectra, but a sodiated glycoside ion $[G+Na]^+$ at *m/z* 185 was only observed for flavonol **17**. Both flavonols **6** and **17** had been reported in Arabidopsis [22].

Although the flavonols **17** and **18** were chromatographically not well resolved, they were distinguished by EICs of the corresponding aglycone fragments (Supplementary figure 41). For flavonol **18**, the remaining fragments were correlated with the structure of Q-R-3-R-7, which had been reported in Arabidopsis before [58].



Putative flavonol	No.	R3'	R3	R7	unclear
K-R-3-G-7	**6**	H	R	G	
K-G-3-R-7	**17**	H	G	R	
Q-R-3-R-7	**18**	OH	R	R	
K+R+G	**22**	H			G, R
K+RG	**25**	H			RG
K+R+G	**30**	H			G, R

Scheme 18: Diglycosylated flavonols $C_{27}H_{30}O_{15}$

Flavonol **25** was assigned as rhamnosyl-glucoside kaempferol based on a $[GR+Na]^+$ fragment and on the high abundance of the $[K–H]^-$ fragment that was reported characteristically for mono-*O*-diglycosides [48], but the glycosylation position could not be designated. Low-abundant flavonols **22** and **30** were only present in protoplast samples and both showed similar spectra, but the glycosylation pattern could not unambiguously be assigned, thus they were both attributed to K+R+G.

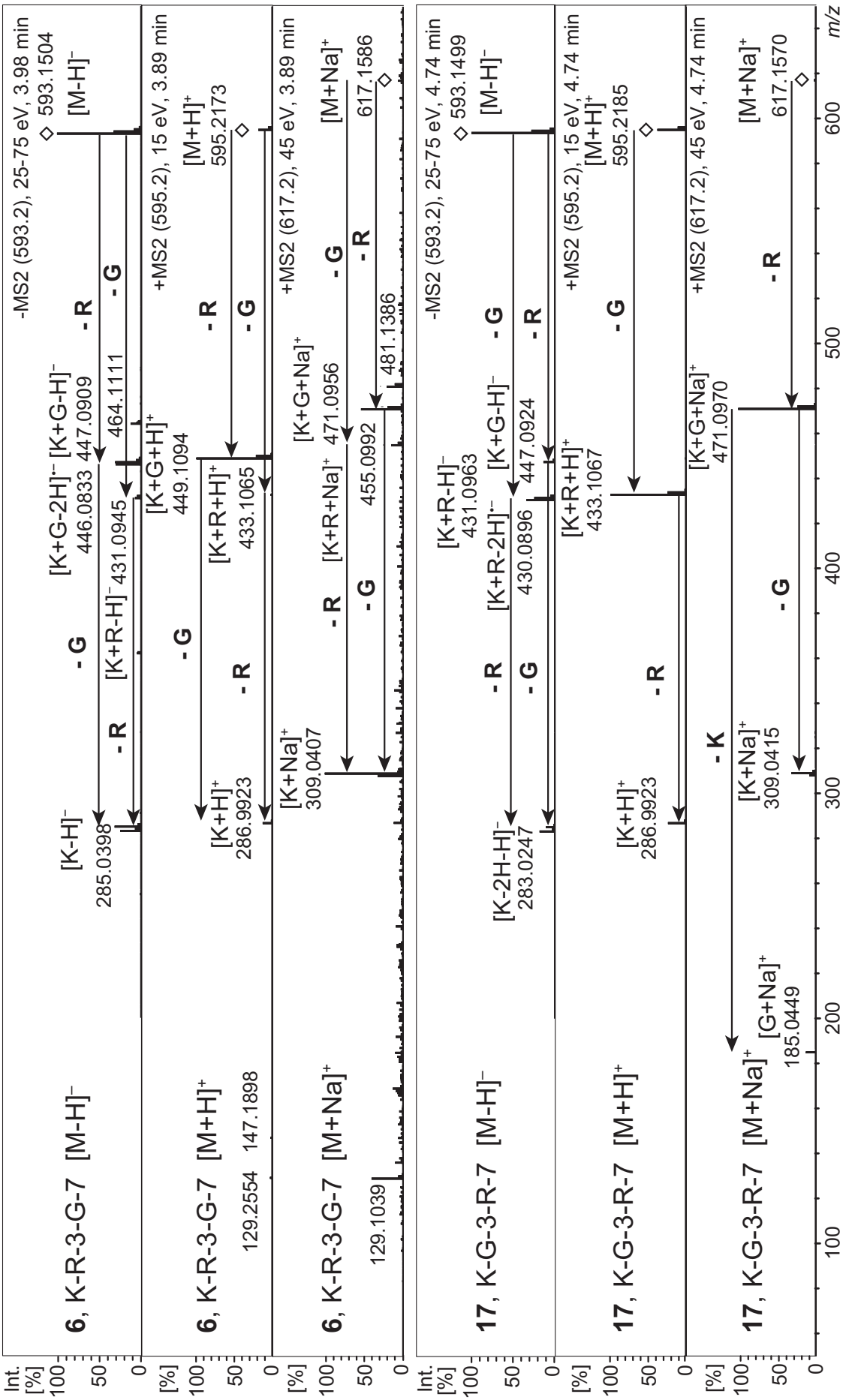
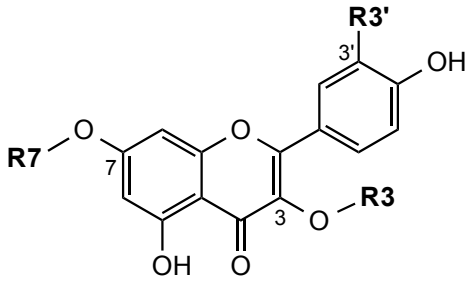


Figure 3: (–)- and (+)-ESI CID MS/MS spectra of diglycosylated flavonols **6** and **17**, putatively assigned as K-R-3-G-7 and K-G-3-R-7.

Five flavonols with formula $C_{27}H_{30}O_{16}$ were detected (Scheme 19), three with a kaempferol core (flavonol **3**, **12**, **26**) and two with a quercetin backbone (flavonol **11**, **20**).



Putative flavonol	No.	R3'	R3	R7	unclear
K-G-3-G-7	**3**	H	G		
Q-G-3-R-7	**11**	OH	G	R	
K-G-3+G	**12**	H	G		G
Q+G+R	**20**	OH			G, R
K+GG	**26**	H			GG

Scheme 19: Diglycosylated flavonols $C_{27}H_{30}O_{16}$

Flavonol **3** was putatively assigned as 3,7-di-*O*-glucoside K-G-3-G-7 based on the (–)-ESI-MS/MS fragment $[C_{23}H_{21}O_{12}]^-$ that corresponded to a neutral loss of $C_4H_8O_4$ from the precursor ion caused by the rupture of a glucoside, and K-G-3-G-7 had been previously detected in Arabidopsis [48]. In (+)-ESI-MS/MS, interference of coeluting fragments corresponding to Q-R-G were observed for the protonated precursor ion, but no sodiated glucoside fragment was detected after fragmentation of $[M+Na]^+$. In flavonols **3**, **11** and **12**, the ions with *m/z* 446, corresponding to $[K+G–2H]^{\bullet-}$ and $[Q+R–2H]^{\bullet-}$ respectively, had been described as indicative for 3-*O*-glycosylation [2], similar to the flavonols **6** and **17** described above.

Flavonol **11** was assigned Q-G-3-R-7 by similarity of its MS/MS spectra that showed a prevalent loss of the 3-*O*-glucoside with spectra from literature and database (e.g. [39], *ReSpect* compound PT211753), due to references in Arabidopsis ([11, 22, 39], and by a missing peak in the *rol1-2 ugt89c1* mutant.

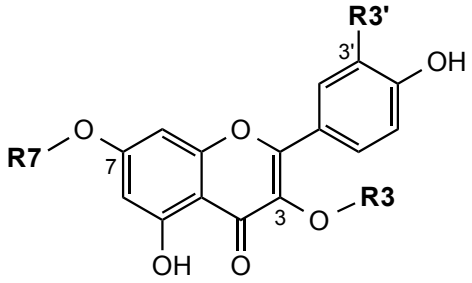
Flavonols **12** and **26** were exclusively detected in *tt7* and *rol1-2 ugt89c1* mutants. Assignment of flavonol **12** was restricted to K-G-3+G because the position of the second glucoside could not be determined. The (–)-ESI-MS/MS spectrum of flavonol **26** showed a neutral loss of a diglucoside (K-GG) corresponding to kaempferol-3-*O*-gentobiose (K-(G1→2G)-3) [48], but this compound was not reported in Arabidopsis so far. Due to low abundance, no MS/MS spectra could be obtained for flavonol **26** in positive ionization mode.

The MS/MS fragmentation spectrum of $[M+H]^+$ of flavonol **20** suggested a 3,7-di-*O*-glycoside, which was supported by the detection of a $[G+Na]^+$, while the (–)-ESI MS/MS spectrum indicated a mono-*O*-diglycoside. However, no sodiated diglycoside fragment

$[GR+Na]^+$ was detected, and no conclusive assignment of flavonol **20** (Q-G-R) could be obtained.

10.5.3 Triglycosylated flavonols

Three kaempferol-hexose-di-rhamnose flavonols with formula $C_{33}H_{40}O_{19}$ were detected in the current Arabidopsis data (Scheme 20).

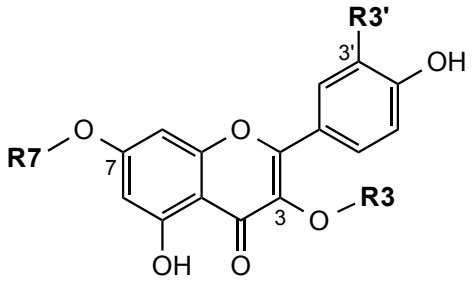


Putative flavonol	No.	R3'	R3	R7	unclear
K-RG-3-R-7	**4**	H	RG	R	
K-R-7+G+R	**16**	H		R	G, R
K-R-7+G+R	**21**	H		R	G, R

Scheme 20: Triglycosylated flavonols $C_{33}H_{40}O_{19}$

The MS/MS spectra of precursors $[M–H]^–$, $[M+H]^+$, $[M+Na]^+$ were very similar: all three flavonols showed a $[GR+Na]^+$ ion indicative for a rhamnosyl-glycosyl diglycosylation and all three flavonols were absent in *rol1-2 ugt89c1*, indicative for 7-*O*-rhamnosylation. Flavonol **4** was suggested as 3-*O*-glucoside, based on the $[K+R–2H]^{\bullet–}$ observed presumably after the homolytic cleavage of the RG-diglycoside at the 3-*O*-position. High similarity of flavonol **4** was observed with kaempferol 3-*O*-rhamnosyl-glucoside 7-*O*-rhamnoside in [27], for which MS^3 [27] or pseudo third-order CID [22] was performed to confirm the loss of the rhamnosyl-hexose. Taken together, the most abundant flavonol **4** was assumed to be K-RG-3-R-7 previously detected in Arabidopsis. Based on references that applied very similar chromatography [59, 73], the retention order could be reproduced for the flavonols **2** < **4** < **11** < **17** < **24**. Therefore, flavonol **4** (K-RG-3-R-7) was postulated to be linked by a [rhamnosyl(1→2)-glucoside] interglycosidic bond.

Flavonols **16** and **21** were rather low abundant. Both isomeric flavonols **16** and **21** were assigned as K-R-7+G+R and are expected to have high structural similarity, but the exact glycosylation pattern could not be determined with the existing data. Additionally, ions which could not be interpreted within the known fragmentation patterns of flavonols were detected for flavonol **21** in (–)-ESI-MS/MS at *m/z* 647, and for flavonol **16** in (+)-ESI-MS/MS at *m/z* 399.



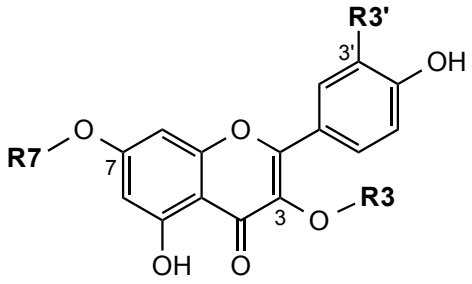
Putative flavonol	No.	R3'	R3	R7	unclear
Q-RG-3-R-7	**2**	OH	RG	R	
K-G-3+G+R	**9**	H	G		G, R
Q-R-3+R+G	**10**	OH	R		R, G
K+G+GR	**13**	H			G, GR
K+G+G+R	**14**	H			G, G, R
K+G+G+R	**15**	H			G, G, R

Scheme 21: Triglycosylated flavonols $C_{33}H_{40}O_{20}$

Of the flavonol triglycosides with formula $C_{33}H_{40}O_{20}$ displayed in Scheme 21, two flavonols **2** and **10** were combinations of a quercetin core with one glycosyl- and two rhamnosyl units. The more abundant flavonol **2** was putatively assigned as Q-RG-3-R-7 supported by the evidence from (–)-ESI-MS/MS (abundant $[Q+R–2H]^{\bullet-}$ indicative for 3-*O*-glycosyl), from sodiated fragments ($[GR+Na]^+$, $[G+Na]^+$), from rhamnosyltransferase mutant (absent in *rol1-2 ugt89c1*), and in accordance with previous reports in Arabidopsis [60]. With reference to a similar chromatography described above for flavonol **4**, the inter-glycosidic linkage of flavonol **2** was assumed as [rhamnosyl(1 → 2)-glucoside] [59, 73]. Flavonol **10** (Q-R-3+R+G) was not rhamnosylated at 7-*O*-position, and 3-*O*-rhamnosylation was assumed based on the high intensity of the $[Q+R–2H]^-$ ion compared to the $[Q+R–H]^-$ ion, although both ions were detected with rather low intensity.

Four kaempferol flavonols with formula $C_{33}H_{40}O_{20}$ were observed. None of them were observed in the rhamnosyltransferase mutant, suggesting 7-*O*-rhamnosyl-modifications, but the assignment was not fully certain due to low abundance of the peaks in Columbia wild type. For flavonols **9**, **13** and **15** precursor ions could only be isolated for MS/MS fragmentation in (–)-ESI, which hampered structure assignment. Fragmentation of flavonol **9** resulted in $[K+G+R–H]^-$ at *m/z* 593 as base peak, suggesting 3-*O*-glycosylation and subsequent losses of a rhamnosyl a further glycosyl unit. Thus, flavonol **9** was assigned as K-G-3+G+R. For flavonol **13**, neutral loss of a rhamnosyl moiety formed the main fragment $[K+G+G–H]^-$ at *m/z* 609, and a small kaempferol ion was detected besides additional unknown ions. Thus, flavonol **13** was proposed K+G+GR. Dissociation of flavonols **14** and **15** lead to (–)-ESI-MS/MS spectra of high similarity. Because MS/MS spectra of K-(GG)-3-R-7 reported in [27] showed no fragment $[K+G+R–H]^-$ and $[K+G–H]^-$, a diglycoside unit was considered unlikely. The competition between both eliminations relates to distinct glycosylation positions in both flavonol **14** and **15**. Based on the available data, no clear

conclusion could be drawn on the attachment of the di-*O*-glycosyl-residues for flavonol **14** and **15** (K+G+G+R).



Putative flavonol	No.	R3'	R3	R7	unclear
K+G+G+G	**1**	H			G, G, G
Q-G-3-R-7+G	**5**	OH	G	R	G
Q-G<u>G</u>-3-R-7	**7**	OH	GG	R	
Q+G+G+R	**8**	OH			G, GR

Scheme 22: Triglycoside flavonols $C_{33}H_{40}O_{21}$

Four flavonols with formula $C_{33}H_{40}O_{21}$ were observed in the current data (Scheme 22).

Flavonol **7** was putatively assigned as Q-G<u>G</u>-3-R-7 due to the absence of the peak in *rol1-2 ugt89c1* indicative for 7-*O*-rhamnosylation, due to an intense sodiated diglycoside fragment $[GG+Na]^+$ detected at *m/z* 347, and due to previously characterized Q-G<u>G</u>-3-R-7 in Arabidopsis [20, 39, 62]. Based on [20] describing the inability of Arabidopsis wild type accession Columbia to form flavonol 3-*O*-gentiobioside 7-*O*-rhamnoside (K-G(1→6)G-3-R-7), the interglycosidic linkage G(1→6)G could be ruled out, leaving G(1→2)G glycosidic linkage as the more reasonable option.

Fragmentation of flavonol **1** led to a subsequent loss of three glucosides revealing a kaempferol aglycone, but no further evidence on the glycan structure could be obtained (K+G+G+G).

Flavonol **5** showed fragmentation of the protonated precursor similar to flavonol **7** with a preferred loss of glycosyl pointing towards 3-*O*-glycosylation, but more pronounced competitive elimination of rhamnosyl and glucosyl moieties in (–)-ESI-MS/MS. Only a $[G+Na]^+$ ion was detected but no sodiated diglycosyl ion, and 7-*O*-rhamnosylation was assigned based on the rhamnosyltransferase mutant. Taken together, flavonol **5** was putatively assigned Q-G-3-R-7+G.

Low abundance of flavonol **8** resulted in poor MS/MS spectra quality in (–)-ESI and also in an uncertain evaluation of the *rol1-2 ugt89c1* mutant, where a small correlating peak was detected. A sodiated rhamnosyl-glucosyl moiety was detected for flavonol **8**, but the strong $[Q+R+H]^+$ fragment after two subsequent losses of glucose in (+)-ESI-MS/MS did not match with this observation. Thus, the assignment of flavonol **8** was appointed as Q+G+G+R.

11 Conclusion

A method for the targeted analysis of flavonols in *Arabidopsis thaliana* had successfully been transferred from HPLC-MS to UHPLC-HR-MS with improvements in sensitivity, chemical formula calculation and analysis speed. Annotations of the flavonols were verified by the help of UV spectra, biological mutants and MS/MS fragmentation. The chemical formulas of the flavonols could accurately be determined, and the flavonol aglycones were assigned on the basis of the corresponding ions in the (+)- and (–)-ESI-MS/MS spectra as well as based on presence or absence of the respective flavonols in the Arabidopsis mutants *tt7*. The glycosylation positions for seven monoglycosides, seven di-*O*-glycosides and three triglycosides were determined based on ESI-MS/MS fragments that were compared to literature, and further confirmation for the flavonol assignments was obtained from the Arabidopsis *rol1-2 ugt89c1* mutant. For a series of six diglycosides and ten triglycosides, only partial information about the glycosylation positions or the nature of glyco-conjugates was obtained that resulted in a lower level of confidence in the interpretation. For some low abundant flavonols, it was not possible to register mass spectra of sufficient quality, or the fragmentation patterns were not conclusive. For further unambiguous determination of the glycosylation positions and the glycan sequence, MS^n experiments might be informative [27, 40].

12 Applications of targeted analysis of flavonols in Arabidopsis

The optimized UHPLC-HR-MS method described in the previous chapter for the targeted analysis of flavonols in *Arabidopsis thaliana* had been successfully applied to several biological studies for the investigation of the complex and diverse functions of flavonols. The results of the biological studies are either published as research articles, or present as manuscripts, and the abstracts are listed in the following section.

12.1.1 7-Rhamnosylated flavonols modulate homeostasis of the plant hormone auxin and affect plant development

Abstract of [21]:

Benjamin M. Kuhn*, Sanae Errafi*, Rahel Bucher, Petre Dobrev, Markus Geisler, Laurent Bigler, Eva Zažímalová, Christoph Ringli, 7-Rhamnosylated flavonols modulate homeostasis of the plant hormone auxin and affect plant development, *Journal of Biological Chemistry*, **2016**, *291*, 10, 5385-5395

* These authors contributed equally to this work

Flavonols are a group of secondary metabolites that affect diverse cellular processes. They are considered putative negative regulators of the transport of the phytohormone auxin, by which they influence auxin distribution and concomitantly take part in the control of plant organ development. Flavonols are accumulating in a large number of glycosidic forms. Whether these have distinct functions and diverse cellular targets is not well understood. The *rol1-2* mutant of *Arabidopsis thaliana* is characterized by a modified flavonol glycosylation profile that is inducing changes in auxin transport and growth defects in shoot tissues. To determine whether specific flavonol glycosides are responsible for these phenotypes, a suppressor screen was performed on the *rol1-2* mutant, resulting in the identification of an allelic series of *UGT89C1*, a gene encoding a flavonol 7-O-rhamnosyltransferase. A detailed analysis revealed that interfering with flavonol rhamnosylation increases the concentration of auxin precursors and auxin metabolites, while auxin transport is not affected. This finding provides an additional level of complexity to the possible ways by which flavonols influence auxin distribution and suggests that flavonol glycosides play an important role in regulating plant development.

12.2 Flavonol-induced changes in PIN2 polarity and auxin transport in the *Arabidopsis thaliana rol1-2* mutant require phosphatase activity

Abstract of [74]:
Benjamin M. Kuhn, Tomasz Nodzyński, Sanae Errafi, Rahel Bucher, Laurent Bigler, Markus Geisler, Jiří Friml, Christoph Ringli, Flavonol-induced changes in PIN2 polarity and auxin transport in the *Arabidopsis thaliana rol1-2* mutant require phosphatase activity, *under revision*, **2016**

The phytohormone auxin is a major determinant and regulatory component important for plant development. Auxin transport between cells is mediated by a complex system of transporters such as PIN and ABCB proteins, and their activity is influenced by phosphatases and kinases. Flavonoids are secondary metabolites that modify auxin accumulation and transport activity. The *Arabidopsis thaliana rol1-2* mutant is affected in the accumulation of flavonols, a subgroup of flavonoids, and is characterized by defects in cell growth and altered auxin transport activity. A new mutation *in ROOTS CURL IN NPA 1* (*RCN1*), encoding a regulatory subunit of the phosphatase PP2A, was identified as a suppressor of *rol1-2. rol1-2 rcn1-3* double mutants show alleviation of the growth defects and a reversion of the auxin transport activity to wild type-like levels. PIN protein localization revealed a basal-to-apical shift of PIN2 in cortical cells of the *rol1-2* mutant, which is reversed in *rol1-2 rcn1-3* to basal localization. *In vivo* analysis of PINOID action, a kinase known to influence PIN protein localization in a PP2A-antagonistic manner, revealed a negative impact of flavonols on PINOID activity. Together, these data indicate that flavonols affect auxin transport activity by modifying the antagonistic kinase/phosphatase equilibrium.

12.3 Multiple functions of the Arabidopsis flavonol synthase FLS1 in flavonol biosynthesis, transcriptional regulation, and cell growth

Abstract of [75]:
Sanae Errafi, Benjamin M. Kuhn, Rahel Bucher, Laurent Bigler, Christoph Ringli, Multiple functions of the Arabidopsis flavonol synthase FLS1 in flavonol biosynthesis, transcriptional regulation, and cell growth, *submitted*, **2016**

Enzymes with well-defined biochemical activities can sometimes have additional functions and show unexpected subcellular distribution. One such enzyme is the *Arabidopsis*

thaliana flavonol synthase FLS1. We found that, besides catalyzing the conversion of dihydroflavonols to flavonols, the last step in flavonol biosynthesis, FLS1 also influences transcriptional activity and cell growth processes. We show that the additional functions of FLS1 strictly depend on its subcellular localization. While an *fls1* mutant shows changes in gene expression compared to the wild type, nuclear FLS1 is required to re-initiate a wild-type expression pattern. Overexpression of *FLS1* in root hairs results in increased root hair elongation, a process that requires nuclear FLS1. Although cytoplasmic FLS1 is sufficient for accumulation of flavonols which then effectively diffuse into the nucleus, nuclear FLS1 is important for the observed enhanced cell growth. The different activities of FLS1 can be genetically separated, resulting in FLS1 variants that can synthesize flavonols but are impaired in the other processes. Together, these results suggest that a single enzyme can have several partly independent functions that might help to integrate input from the plant's metabolic state onto different processes controlling cell growth and development.

12.4 Light receptors modify flavonol accumulation and flavonol-induced changes in plant cell development

Abstract of [76]:

Marie–Therese Abdou*, Rahel Bucher*, Laurent Bigler, Christoph Ringli, Light receptors modify flavonol accumulation and flavonol-induced changes in plant cell development, *in preparation*, **2016**

* These authors contributed equally to this work

Flavonoids are a group of plant secondary metabolites that serve multiple functions including protection from UV radiation. Flavonols and anthocyanins are abundant components of the group of flavonoids and are glycosylated, resulting in a large number of diverse compounds with distinct properties and functions. Not only the absolute amount of these compounds can vary, but also the degree of glycosylation. The *Arabidopsis thaliana* mutant *rol1-2* shows an altered flavonol glycosylation profile and previous analyses have shown that this alteration correlates with developmental defects found in *rol1-2* seedlings. Here, we characterized the influence of light and light perception on the development of the *rol1-2* growth defects. Exposing this mutant to monochromatic light modulates the flavonol glycosylation and reduces the mutant growth phenotypes to different degrees. Both forward and reverse genetic analyses revealed that mutating the red-light receptor *phyB* completely suppresses the *rol1-2* phenotype while a *cry1* blue-light receptor mutation

partially alleviates the defects observed in *rol1-2*. Thus, alterations in light irradiation and the activity of light receptors modifies flavonol glycosylation and in this way can alter plant development.

13 Experimental procedures flavonol analysis in Arabidopsis

Sample preparation including harvest, seedling separation into shoot and root, extraction and centrifugation was done within the collaborating group of Prof. Christoph Ringli, Department of Plant and Microbial Biology, UZH according to [21]. Briefly, Arabidopsis seedlings were grown in vertical orientation on standard half-strength *Murashige & Skoog* medium (*Sigma–Aldrich*, Buchs, Switzerland) for 6 days in a regime with 16 h light / 8 h dark at 22 °C. One hundred seedlings were separated into root and shoot tissue just above the root transition zone, and frozen in liquid nitrogen. Samples were lyophilized, weighed and extracted with 500 µl extraction solvent (80% [v/v] MeOH in ultrapure water) overnight at 4 °C. Extracts were macerated with pestles, vortexed and after centrifugation 400 µl supernatant was transferred into fresh tubes. After solvent evaporation in speed vac, pellets were dissolved in 100 µl fresh extraction solvent and stored at 20 °C upon mass spectrometry analysis.

13.1 Chemicals and reagents

Mass spectrometry grade acetonitrile (MeCN) and formic acid (99.995%) were purchased from Fluka (Buchs, Switzerland). Water was obtained from a Millipore high purity water dispenser (Billerica, MA).

13.2 UHPLC-DAD-HR-MS analysis

UHPLC-HR-MS analyses were performed on an *Acquity UPLC* (*Waters*, Baden-Dättwil, Switzerland) connected to a maXis high-resolution Q-TOF mass spectrometer (*Bruker Daltonics*, Fällanden, Switzerland), driven by *Compass Hystar* 3.2 and *Compass microTOFcontrol* 1.3 software. Flavonols were separated on an *Acquity BEH C18* reversed phase column (1.7 mm, 2.1 x 100 mm, guard column 2.1 x 5 mm, *Waters*, Baden-Dättwil, Switzerland) kept at 30 °C. Mobile phases consisted of water/MeCN containing 0.1% [v/v] formic acid. Gradient elution at flow rate 0.45 ml min^{-1} was performed from 5% to 30% MeCN in 8 min, followed by a gradient 30–99.5% MeCN in 1 min, a washing step at 99.5% for 1 min, and re-equilibration at 5% MeCN for 2 min, leading to a final runtime of 12 min. A volume of 2 µl sample was injected, followed by 600 µl weak wash (H_2O : MeCN 95:5 [v/v]) and 400 µl strong wash (H_2O : MeCN 3:7 [v/v]). Samples were stored in auto sampler at 10 °C.

UV spectra were recorded on a diode array detector from 200–600 nm with 1.2 nm resolution and 20 points s^{-1} sampling rate.

The mass spectrometer was operated in negative electrospray ionization mode at 3500 V capillary voltage, –500 V endplate offset, with a N_2 nebulizer pressure of 4 bar and dry gas flow of 9.5 l min^{-1} at 205 °C. MS acquisitions were performed in the full scan mode in the mass range from mass-to-charge ratio (*m/z*) 50 to 1200 at 25,000 resolution (full width at half maximum) and 2 scans per second. Masses were calibrated prior to analysis with a 2 mM solution of sodium formiate over a mass range of *m/z* 249 to 1337. Samples were acquired using dead volume calibration with sodium formiate solution.

For CID MS^2 experiments, precursor ions $[M–H]^-$, $[M+H]^+$ and $[M+Na]^+$ were isolated with 6 or 4 Da isolation width in negative and positive mode respectively. Negative ion spectra were recorded at 6 Hz with a ramped collision energy depending on the flavonol from 25–75 eV, positive ion spectra at 2 Hz with 15 eV ($[M+H]^+$) and 45 eV ($[M+Na]^+$) with 4000 V capillary voltage and 400 Vpp collision RF. For production of sodiated adducts, aqueous sodium formiate solution (2 mM) was added via post-column infusion at 3 µl min^{-1} flow rate.

13.3 Data processing

Raw data were calibrated in *DataAnalyis* 4.0 (*Bruker Daltonics*, Fällanden, Switzerland) from *m/z* 248–860 with search range 0.05 *m/z*, 1000 intensity threshold and HPC mode, and the overall deviation was manually checked to be below 0.3 ppm.

Peaks of extracted ion chromatograms with flavonol *m/z* $[M–H]^-$ ± 0.01 were integrated in *QuantAnalysis* 4.2 (*Bruker Daltonics*), and the automatic integrations were inspected manually. The area under the curve was used for relative quantification for comparison between different mutants or conditions.

14 References

[1] Andersen, Ø. M., Markham, K. R., *Flavonoids - Chemistry, Biochemistry and Applications*, CRC Press, Boca Raton, FL, **2006**

[2] Vukics, V., Guttman, A., *Mass Spectrom. Rev.,* **2010**, *29*, 1, 1-16

[3] de Villiers, A., Venter, P., Pasch, H., *J. Chromatogr. A,* **2016**, *1430*, 16-78

[4] Taylor, L. P., Grotewold, E., *Curr. Opin. Plant Biol.,* **2005**, *8*, 3, 317-323

[5] Agati, G., Azzarello, E., Pollastri, S., Tattini, M., *Plant Sci.,* **2012**, *196*, 67-76

[6] Buer, C. S., Muday, G. K., *Plant Cell,* **2004**, *16*, 5, 1191-1205

[7] Peer, W. A., Blakeslee, J. J., Yang, H., Murphy, A. S., *Molecular plant,* **2011**, *4*, 3, 487-504

[8] Nakabayashi, R., Saito, K., *Anal. Bioanal. Chem.,* **2013**, *405*, 15, 5005-5011

[9] Koornneef, M., Meinke, D., *Plant J.,* **2010**, *61*, 6, 909-921

[10] Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., Huala, E., *Genesis,* **2015**, *53*, 8, 474-485

[11] Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., Fernie, A. R., *Plant Physiol. Biochem.,* **2013**, *72*, 21–34

[12] Davies, K. M., Schwinn, K. E., in *Flavonoids - Chemistry, Biochemistry and Applications* (Eds.: Andersen, Ø. M., Markham, K. R.), CRC Press, Boca Raton, FL, **2006**, pp. 143-218.

[13] Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., Caboche, M., *Annu. Rev. Plant Biol.,* **2006**, *57*, 405-430

[14] Koornneef, M., *Arab. Inf. Serv,* **1990**, *19*, 113-115

[15] Böttcher, C., von Roepenack-Lahaye, E., Schmidt, J., Schmotz, C., Neumann, S., Scheel, D., Clemens, S., *Plant Physiol.,* **2008**, *147*, 4, 2107-2120

[16] Routaboul, J. M., Kerhoas, L., Debeaujon, I., Pourcel, L., Caboche, M., Einhorn, J., Lepiniec, L., *Planta,* **2006**, *224*, 1, 96-107

[17] Kuhn, B. M., Geisler, M., Bigler, L., Ringli, C., *Plant Physiol.,* **2011**, *156*, 2, 585-595

[18] Diet, A., Link, B., Seifert, G. J., Schellenberg, B., Wagner, U., Pauly, M., Reiter, W. D., Ringli, C., *Plant Cell,* **2006**, *18*, 7, 1630-1641

[19] Ringli, C., Bigler, L., Kuhn, B. M., Leiber, R.-M., Diet, A., Santelia, D., Frey, B., Pollmann, S., Klein, M., *Plant Cell,* **2008**, *20*, 6, 1470-1481

[20] Ishihara, H., Tohge, T., Viehover, P., Fernie, A. R., Weisshaar, B., Stracke, R., *J. Exp. Bot.,* **2016**, *67*, 5, 1505-1517

[21] Kuhn, B. M., Errafi, S., Bucher, R., Dobrev, P., Geisler, M., Bigler, L., Zažímalová, E., Ringli, C., *J. Biol Chem.,* **2016**, *291*, 10, 5385-5395

[22] Kerhoas, L., Aouak, D., Cingoz, A., Routaboul, J.-M., Lepiniec, L., Einhorn, J., Birlirakis, N., *J. Agric. Food Chem.,* **2006**, *54*, 18, 6603-6612

[23] Veit, M., Pauli, G. F., *J. Nat. Prod.,* **1999**, *62*, 1301-1303

[24] van der Hooft, J. J., Mihaleva, V., de Vos, R. C., Bino, R. J., Vervoort, J., *Magn. Reson. Chem.,* **2011**, *49 Suppl 1*, S55-60

[25] Sumner, L. W., Mendes, P., Dixon, R. A., *Phytochemistry,* **2003**, *62*, 6, 817-836

[26] Abad-Garcia, B., Berrueta, L. A., Garmon-Lobato, S., Gallo, B., Vicente, F., *J. Chromatogr. A,* **2009**, *1216*, 28, 5398-5415

[27] Stobiecki, M., Skirycz, A., Kerhoas, L., Kachlicki, P., Muth, D., Einhorn, J., Mueller-Roeber, B., *Metabolomics,* **2006**, *2*, 4, 197-219

[28] Cuyckens, F., Shahat, A. A., Pieters, L., Claeys, M., *J. Mass Spectrom.,* **2002**, *37*, 12, 1272-1279

[29] Rauha, J.-P., Vuorela, H., Kostiainen, R., *J. Mass Spectrom.,* **2001**, *36*, 12, 1269-1280

[30] March, R., Brodbelt, J., *J. Mass Spectrom.,* **2008**, *43*, 12, 1581-1617

[31] Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., Akiyama, K., Sakurai, T., Matsuda, F., Aoki, T., Hirai, M. Y., Saito, K., *Phytochemistry,* **2012**, *82*, 38-45

[32] Horai, H.*, et al.*, *J. Mass Spectrom.,* **2010**, *45*, 7, 703-714

[33] Rothwell, J. A., Perez-Jimenez, J., Neveu, V., Medina-Remon, A., M'Hiri, N., Garcia-Lobato, P., Manach, C., Knox, C., Eisner, R., Wishart, D. S., Scalbert, A., *Database (Oxford),* **2013**, *2013*, bat070

[34] Cuyckens, F., Claeys, M., *J. Mass Spectrom.,* **2004**, *39*, 1, 1-15

[35] Fabre, N., Rustan, I., de Hoffmann, E., Quetin-Leclercq, J., *J. Am. Soc. Mass Spectrom.,* **2001**, *12*, 707-715

[36] Brazier-Hicks, M., Evans, K. M., Gershater, M. C., Puschmann, H., Steel, P. G., Edwards, R., *J. Biol. Chem.,* **2009**, *284*, 27, 17926-17934

[37] Yun, C. S., Yamamoto, T., Nozawa, A., Tozawa, Y., *Biosci. Biotechnol. Biochem.,* **2008**, *72*, 4, 968-973

[38] Falcone Ferreyra, M. L., Emiliani, J., Rodriguez, E. J., Campos-Bermudez, V. A., Grotewold, E., Casati, P., *Plant Physiol.,* **2015**, *169*, 2, 1090-1107

[39] Kachlicki, P., Einhorn, J., Muth, D., Kerhoas, L., Stobiecki, M., *J. Mass Spectrom.,* **2008**, *43*, 5, 572-586

[40] Ablajan, K., Abliz, Z., Shang, X. Y., He, J. M., Zhang, R. P., Shi, J. G., *J. Mass Spectrom.,* **2006**, *41*, 3, 352-360

[41] Hvattum, E., Ekeberg, D., *J. Mass Spectrom.,* **2003**, *38*, 1, 43-49

[42] Wojakowska, A., Piasecka, A., Garcia-Lopez, P. M., Zamora-Natera, F., Krajewski, P., Marczak, L., Kachlicki, P., Stobiecki, M., *Phytochemistry,* **2013**, *92*, 71-86

[43] Stobiecki, M., Kachlicki, P., Wojakowska, A., Marczak, Ł., *Phytochemistry Letters,* **2015**, *11*, 358-367

[44] Ma, Y. L., Vedernikova, I., Van den Heuvel, H., Claeys, M., *J. Am. Soc. Mass Spectrom.,* **2000**, *11*, 2, 136-144

[45] Abad-Garcia, B., Garmon-Lobato, S., Berrueta, L. A., Gallo, B., Vicente, F., *J. Mass Spectrom.,* **2009**, *44*, 7, 1017-1025

[46] Cuyckens, F., Ma, Y. L., Pocsfalvi, G., Claeysi, M., *Analusis,* **2000**, *28*, 10, 888-895

[47] Shi, P., He, Q., Song, Y., Qu, H., Cheng, Y., *Anal. Chim. Acta,* **2007**, *598*, 1, 110-118

[48] Ferreres, F., Llorach, R., Gil-Izquierdo, A., *J. Mass Spectrom.,* **2004**, *39*, 3, 312-321

[49] Kalili, K. M., Vestner, J., Stander, M. A., de Villiers, A., *Anal. Chem.,* **2013**, *85*, 19, 9107-9115

[50] Montero, L., Herrero, M., Prodanov, M., Ibanez, E., Cifuentes, A., *Anal. Bioanal. Chem.,* **2013**, *405*, 13, 4627-4638

[51] Montero, L., Herrero, M., Ibáñez, E., Cifuentes, A., *J. Chromatogr. A,* **2013**, *1313*, 275-283

[52] Wolfender, J. L., Marti, G., Thomas, A., Bertrand, S., *J. Chromatogr. A,* **2015**, *1382*, 136-164

[53] Groessl, M., Azzollini, A., Eugster, P. J., Plet, B., Wolfender, J. L., Knochenmuss, R., *Chimia,* **2014**, *68*, 3, 135-139

[54] Gonzales, G. B., Raes, K., Coelus, S., Struijs, K., Smagghe, G., Van Camp, J., *J. Chromatogr. A,* **2014**, *1323*, 39-48

[55] Yassin, G. H., Grun, C., Koek, J. H., Assaf, K. I., Kuhnert, N., *J. Mass Spectrom.,* **2014**, *49*, 11, 1086-1095

[56] Silvestro, L., Tarcomnicu, I., Dulea, C., Attili, N. R., Ciuca, V., Peru, D., Rizea Savu, S., *Anal. Bioanal. Chem.,* **2013**, *405*, 25, 8295-8310

[57] Santos-Buelga, C., García-Viguera, C., Tomas-Barberan, F. A., in *Methods in Polyphenol Analysis,* (Eds.: Santos-Buelga, C., Williamson, G.), TheRoyal Society of Chemistry, Cambridge, **2003**, pp. 92–127.

[58] Sugimoto, T., Bamba, T., Izumi, Y., Nomura, H., Shiina, T., Fukusaki, E., *J. Sep. Sci.,* **2011**, *34*, 24, 3587-3596

[59] Tohge, T., Yonekura-Sakakibara, K., Niida, R., Watanabe-Takahashi, A., Saito, K., *Pure Appl. Chem.,* **2007**, *79*, 4

[60] Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., Watanabe-Takahashi, A., Inoue, E., Saito, K., *Plant Cell,* **2008**, *20*, 8, 2160-2176

[61] Zhang, J., Sun, X., Zhang, Z., Ni, Y., Zhang, Q., Liang, X., Xiao, H., Chen, J., Tokuhisa, J. G., *Phytochemistry,* **2011**, *72*, 14-15, 1767-1778

[62] Matsuda, F., Hirai, M. Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N. J., Sakurai, T., Shimada, Y., Saito, K., *Plant Physiol.,* **2009**, *152*, 2, 566-578

[63] Dugo, P., Donato, P., Cacciola, F., Germano, M. P., Rapisarda, A., Mondello, L., *J. Sep. Sci.,* **2009**, *32*, 21, 3627-3634

[64] Kim, S. M., Kang, S. W., Jeon, J. S., Jung, Y. J., Kim, C. Y., Pan, C. H., Um, B. H., *Biomed. Chromatogr.,* **2012**, *26*, 2, 199-207

[65] Ding, Y., Dou, D., Guo, Y., Li, Q., *Pharmacogn. Mag.,* **2014**, *10*, 40, 435-440

[66] Kong, N. N., Fang, S. T., Liu, Y., Wang, J. H., Yang, C. Y., Xia, C. H., *Nat. Prod. Res.,* **2014**, *28*, 12, 928-931

[67] An, H., Wang, H., Lan, Y., Hashi, Y., Chen, S., *J. Pharm. Biomed. Anal.,* **2013**, *85*, 295-304

[68] Aljancic, I. S., Vuckovic, I., Jadranin, M., Pesic, M., Dordevic, I., Podolski-Renic, A., Stojkovic, S., Menkovic, N., Vajs, V. E., Milosavljevic, S. M., *Phytochemistry,* **2014**, *98*, 190-196

[69] Mikulic-Petkovsek, M., Slatnar, A., Stampar, F., Veberic, R., *Food Chem.,* **2012**, *135*, 4, 2138-2146

[70] Sientzoff, P., Hubert, J., Janin, C., Voutquenne-Nazabadioko, L., Renault, J. H., Nuzillard, J. M., Harakat, D., Magid, A. A., *Molecules,* **2015**, *20*, 8, 14970-14984

[71] Refaat, J., Samy, M. N., Desoukey, S. Y., Ramadan, M. A., Sugimoto, S., Matsunami, K., Kamel, M. S., *Med. Chem. Res.,* **2015**, *24*, 7, 2939-2949

[72] Slimestad, R., Andersen, Ø. M., Francis, G. W., Marston, A., Hostettmann, K., *Phytochemistry,* **1995**, *40*, 5, 1537–1542

[73] Stracke, R., Jahns, O., Keck, M., Tohge, T., Niehaus, K., Fernie, A. R., Weisshaar, B., *New Phytol.,* **2010**, *188*, 4, 985-1000

[74] Kuhn, B. M., Nodzyński, T., Errafi, S., Bucher, R., Bigler, L., Geisler, M., Friml, J., Ringli, C., Flavonol-induced changes in PIN2 polarity and auxin transport in the *Arabidopsis thaliana rol1-2* mutant require phosphatase activity, *under revision,* **2016**

[75] Errafi, S., Kuhn, B. M., Bucher, R., Bigler, L., Ringli, C., Multiple functions of the *Arabidopsis* flavonol synthase FLS1 in flavonol biosynthesis, gene expression, and cell growth, *submitted*, **2016**

[76] Abdou, M.-T., Bucher, R., Bigler, L., Ringli, C., Light receptors modify flavonol accumulation and flavonol–induced changes in plant cell development, *in preparation*, **2016**

— Appendix —

Abbreviations

(–)-ESI	Negative ionization mode ESI
(+)-ESI	Positive ionization mode ESI
$[M–H]^-$	Deprotonated molecule ion
$[M+H]^+$	Protonated molecule ion
ABA	Abscisic acid
ABC transporter	Adenosine triphosphate-binding cassette transporter
ANOVA	Analysis of variance
APCI	Atmospheric pressure chemical ionization
ATP	Adenosin triphosphate
B	Barley
BG8	Transgenic line of barley, high expression of *Lr34*
BG9	Transgenic line of barley, very high expression of *Lr34*
BPC	Base peak chromatogram
CAS	Chemical Abstracts Service
CE	Collision energy
Cer	Ceramide
CID	Collision induced dissociation
Ctrl	Control
DESI	Desorption electrospray
DGDG	Digalactosyl diacylglycerol
DIMS	Direct infusion mass spectrometry
EI	Electron impact
EIC	Extracted ion chromatogram
ESI	Electrospray ionization
ETI	Effector triggered immunity
Exp	Experiment
FA	Fatty acid
FC	Fold change
FID	Flame ionization detector
FLS	Flavonol synthase
G	Glucosyl
GC	Gas chromatography
GC-MS	Gas chromatography mass spectrometry
GLP	Germin-like protein
GlyCer	Glucosyl ceramide
H	Hydroponic growth condition
HILIC	Hydrophobic interaction chromatography

HPLC	High performance liquid chromatography
HR	High resolution
I	Isorhamnetin (chapter 3)
I	Pathogen infected growth condition (chapter 2)
JA	Jasmonic acid
K	Kaempferol
LC	Liquid chromatography
LC-MS	Liquid chromatography mass spectrometry
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LIT	Linear ion trap
Lr34	Leaf rust resistance gene *Lr34*
lysoPC	Phosphatidyl choline having lost one FA chain
lysoPE	Phosphatidyl ethanolamine having lost one FA chain
M	Mock infected growth condition
m/z	Mass-to-charge ratio
MALDI	Matrix-assisted laser desorption ionization
MAMP	Microbe associated molecular patterns
MGDG	Monogalactosyl diacylglycerol
MPIMP	Max Planck Institute of Molecular Plant Physiology
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MS^n	Multiple stage mass spectrometry
MTBE	Methyl-tert-buthylether
MW	Molecular weight
N	Non-infected growth condition
NL	Neutral loss
NMR	Nuclear magnetic resonance
p-values	Probability values
PAL	Phenylalanine ammonia lyase
PAMP	Pathogen associated molecular patterns
PC	Phosphatidyl choline
PCA	Principal component analysis
PE	Phosphatidyl ethanolamine
PG	Phosphatidyl glycerol
PI	Phosphatidyl inositol
PS	Phosphatidyl serine
PTI	PAMP-triggered immunity
Q	Quercetin
QC	Quality control
QLIT	Quadrupole - linear ion trap
QqQ	Triple quadrupol

QRL	Quantitative resistance loci
qRT-PCR	Qualitative real-time polymerase chain reaction
QTOF	Quadrupole - time of flight
R	Rhamnosyl (chapter 3)
R	Rice (chapter 2)
R-genes	Resistance genes
RI	Retention time indices
rol1–2 ugt	Mutant unable to synthesize 7–rhamnosyl flavonols
rol1–2	Repressor of *lrx1*
ROS	Reactive oxygen species
RP	Reversed phase
RT	Retention time
SA	Salicylic acid
Sib	Sister line
SIMS	Secondary ion mass spectrometry
SQDG	Sulfoquinovosyl diacylglycerol
TAG	Triacylglycerol
TIC	Total ion chromatogram
TOF	Time-of-flight
tt4	*Transparent testa* mutant *tt4*, deficient in chalcone synthase
tt7	*Transparent testa* mutant *tt7*, defect in flavonol 3'-hydroxylase
UHPLC	Ultra-high performance liquid chromatography
UV	Ultra violet
UZH	University of Zurich
W	Wheat
Wt	Wild type
8	Transgenic line of rice, low expression of *Lr34*
19	Transgenic line of rice, high expression of *Lr34*

Curriculum vitae

Name:	Rahel Ruth Bucher
Date of Birth:	19.01.1984
Origin:	Altstätten SG, CH

Education

Apr 2012 – present	**PhD Student,** Department of Chemistry, University of Zurich, Member of Graduate School of Chemical and Molecular Sciences Zürich, CH - Thesis "Mass-spectrometry based metabolomic approaches to characterize plant pathogen resistance in cereals and the flavonols modifying development of *Arabidopsis thaliana*" Supervision: PD Dr. Laurent Bigler, Prof. Dr. Stefan Bienz, Prof. Dr. Beat Keller
Mar 2015 – Oct 2015	**Visiting PhD Student**, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, DE
Feb 2010 – Oct 2011	**Master of Sciences in Life Science,** Molecular Technologies, University of Applied Sciences of Northwestern Switzerland, Basel CH - Thesis F. Hoffmann – La Roche "Metabolic soft spot analysis in early drug discovery by MassMetaSite in comparison to manual interpretation of MS/MS spectra and NMR spectroscopy"
Sept 2003 – Dec 2006	**Undergraduate Study of Chemistry Chemist FH**, major Biological Chemistry, Zurich University of Applied Sciences, Winterthur CH - Thesis "Purification of green fluorescent protein (GFPuv-his) using metallchelate-membrane chromatography"
Aug 2002 – Jul 2003	**Federal Vocational Baccalaureate**, technical specialization, Gewerbliche Berufsschule, St.Gallen, CH
Aug 1999 – Jul 2002	**Apprenticeship Textile Laboratory Assistant**, Bezema AG, Montlingen CH

Experience

Jan 2012 – Apr 2012	**Labor Specialist,** Drug Metabolism Group, F. Hoffmann – LaRoche, Basel CH - Mass spectrometry applications - Transfer of glutathione reactive metabolite screening assay

Jul 2010 – Aug 2010	**Intern,** Drug Metabolism Group, F. Hoffmann – La Roche, Basel CH - Mass spectrometry applications, metabolite trapping
Nov 2007 – Jan 2010	**Technical Assistant**, Institute for Biopharmaceutical Research IBR Inc., Matzingen CH - Method development, GLP validation of HPLC and GC methods
Apr – Jul 2007	**Intern**, Egret Pharma Ltd., Shanghai CN - Pharmacological and toxicological in vitro assays
Jul – Sept 2005	**Intern**, Microsynth AG, Balgach CH - DNA-sequencing and -synthesis
Jul 2002 – Oct 2002	**Textile Laboratory Assistant**, Bezema AG, Montlingen, CH

Scientific communications and collaborations

Journal articles

Chauhan H., Boni R., Bucher R., Kuhn B., Buchmann G., Sucher J., Selter L. L., Hensel G., Kumlehn J., Bigler L., Glauser G., Wicker T., Krattinger S. G., Keller B., The wheat resistance gene *Lr34* results in the constitutive induction of multiple defense pathways in transgenic barley, *Plant J.*, **2015**, *84*, 1, 202–215, 10.1111/tpj.13001

Kuhn B. M., Errafi S., Bucher R., Dobrev P., Geisler M., Bigler L., Zažímalová E., Ringli C., 7-Rhamnosylated flavonols modulate homeostasis of the plant hormone auxin and affect plant development, *J Biol Chem.,* **2016**, *291*, 10, 5385-5395

Kuhn B. M., Nodzyński T., Errafi S., Bucher R., Bigler L., Geisler M., Friml J., Ringli C., Flavonol-induced changes in PIN2 polarity and auxin transport in the *Arabidopsis thaliana rol1-2* mutant require phosphatase activity, *under revision*, **2016**

Errafi S., Kuhn B. M., Bucher R., Bigler L., Ringli C., Multiple functions of the Arabidopsis flavonol synthase FLS1 in flavonol biosynthesis, transcriptional regulation, and cell growth, *submitted*, **2016**

Abdou M–T., Bucher R., Bigler L., Ringli C., Light receptors modify flavonol accumulation and flavonol-induced changes in plant cell development, *in preparation*, **2016**

Poster communications

1.–4.07.2013	Metabolomics Society Conference, Glasgow, GB, “Analysing durable resistance processes in cereals by metabolic fingerprinting”
10.10.2013	26th Meeting of the Israeli Society for Mass Spectrometry, Rehovot, IL, “Analyzing durable antifungal resistance processes in cereals by metabolic fingerprinting”
10.12.2013	Annual Advances in Swiss Research on the Chemical Ecology of Plant Defense, Neuchatel, CH, “Analysing durable anti-fungal resistance processes in cereals by metabolic fingerprinting”
17.–21.03.2014	EMBO Practical Course on Metabolomics Bioinformatics for Life Scientists, European Bioinformatics Institute, Hinxton, GB, “Analysing durable resistance processes in cereals by metabolic fingerprinting”

24–29.08.2014 International Mass Spectrometry Conference, Geneva, CH, “Analysing durable resistance processes in cereals by metabolic fingerprinting”

08.03.2015 27th Meeting of the Israeli Society for Mass Spectrometry, Rehovot, IL, “Analyzing durable antifungal resistance processes in cereals by metabolic fingerprinting”

30.08.–3.09.2015 Botanikertagung 2015, Freising, Deutschland, “Investigating the metabolome of Poaceae plants displaying a durable anti-fungal resistance phenotype”

Oral communications

03.06.2014 Doktorandentag, Department of Chemistry UZH, Au, CH, “Analyzing durable anti-fungal resistance processes in cereals by metabolic fingerprinting using LC-MS”

03.06.2014 Retreat CMSZH Graduate School UZH, Zuoz, CH, “Durable anti-fungal resistance in cereals analyzed by LC-MS metabolomics”

Grants

Doc.Mobility No. 158657, 01.03.2015 – 31.08.2015

Title: Metabolomics as part of a multi-disciplinary approach to unravel the molecular function of durable disease resistance in cereals

Research institution: Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, DE

Collaborations

Collaboration with the research group of Prof. Dr. Beat Keller, Department of Plant and Microbial Biology, University of Zurich, CH: with Dr. Benjamin Kuhn for establishing hydroponic growth of barley plants; with Dr. Harsh Chauhan for barley plant growth on soil, harvesting, homogenizations and extractions, with Rainer Böni for growth of barley plants on hydroponics.

Metabolomic profiling of *Lr34* resistance in barley, rice and wheat by LC- and GC-MS; with Dr. Simon Krattinger for design of experiment, with Rainer Böni for growth and leaf rust infection of barley plants, with Justine Sucher for establishing hydroponic growth of rice plants and rice blast infection of rice plants, with Jyoti Singla for field-grown wheat plants.

Collaboration with the research group of Prof. Dr. Dr. h.c. Lothar Willmitzer, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, DE:

Metabolomic profiling of *Lr34* resistance in barley, rice and wheat by LC- and GC-MS; with Dr. Dirk Steinhauser for design of experiment, with Änne Michaelis for LC-MS analysis, with Gudrun Wolther for GC-MS analysis, with Dr. Asdrubal Burgos for lipid identification and with Dr. Daniel Veyel for data processing.

Collaboration with the research group of Prof. Dr. Christoph Ringli, Department of Plant and Microbial Biology, University of Zurich, CH:
Flavonol structure assignment and Applications of targeted analysis of flavonols in Arabidopsis; with Prof. Christoph Ringli, Sanae Errafi and Marie-Therese Abdou for the generation of Arabidopsis mutant lines and plant growth.

Acknowledgement

I would like to thank my supervisor PD Dr. Laurent Bigler for giving me the opportunity to perform my PhD, and for providing the exciting facility including all this wonderful mass spectrometers. I thank Prof. Dr. Stefan Bienz for being the responsible faculty member of my PhD committee. I thank Prof. Dr. Beat Keller for his support and the opportunity to work on the exciting and challenging *Lr34* project. I thank all members of my PhD committee for their good will and their precious time to read my manuscript and to improve this thesis, and for enabling me to independently find my way through the world of metabolomics.

I am grateful for financial support of the University of Zurich, the Swiss National Foundation, the Chemical and Molecular Sciences Zürich graduate school, and the laboratories of PD Dr. Laurent Bigler, Prof. Dr. Beat Keller, Prof. Dr. Enrico Martinoia and Prof. Dr. Stefan Bienz.

I thank Urs Stalder for his patient help in the lab, and my colleague Victor and all the members of the Bigler+Bienz group for the small breaks.

I thank Prof. Christoph Ringli for the productive and enjoyable collaboration on Arabidopsis flavonols, and also Sanae Errafi and Marie-Therese Abdou for the careful preparation of the samples. I thank the collaboration partners Harsh, Benjamin, Rainer, Justine, Johannes and Simon for the biological experiments and discussions.

I especially want to thank Dr. Daniel Veyel for the very valuable discussions on plant metabolomics during and after my time at the Max Planck Institute of Molecular Plant Physiology. I thank Dr. Alexander Erban for his expert instructions of *TagFinder*, Dr. Asdrubal Burgos for his help with lipid profiling, Ola, Sylwia, Hezi and Yariv for discussions. I thank Prof. Dr. Dr. Lothar Willmitzer for supporting my visit at the MPIMP and Änne Michaelis and Gudrun Wolther for their technical and moral support.

I warmly thank Prof. Dr. Uwe Pieles for his encouraging input and important input for grant writing. I further want to acknowledge people that inspired me on my way, namely Dr. David Fischer, Dr. Julien Boccard and colleagues I met at courses of IMSS and EMBO.

I thank my friends that gave me perspective beyond graduation, namely Jun, WG Nansenstrasse, the Life Science Zürich Young Scientist Network and my dear friends at Casteldoro.

Von ganzem Herzen bedanke ich mich bei Ismael, der mir während dieser ganzen anspruchsvollen Zeit liebevoll zur Seite stand und mich fortwährend aufmerksam unterstützte.

Ganz herzlichen Dank für die kräftige und unermüdliche Unterstützung gebührt Hedy+Jörg, Tabitha und Jean. Ich bin euch sehr dankbar für euren aufmunternden Beistand in den schwierigen Situationen der Doktorarbeit, von Anfang bis Schluss und darüber hinaus.

Supplementary information

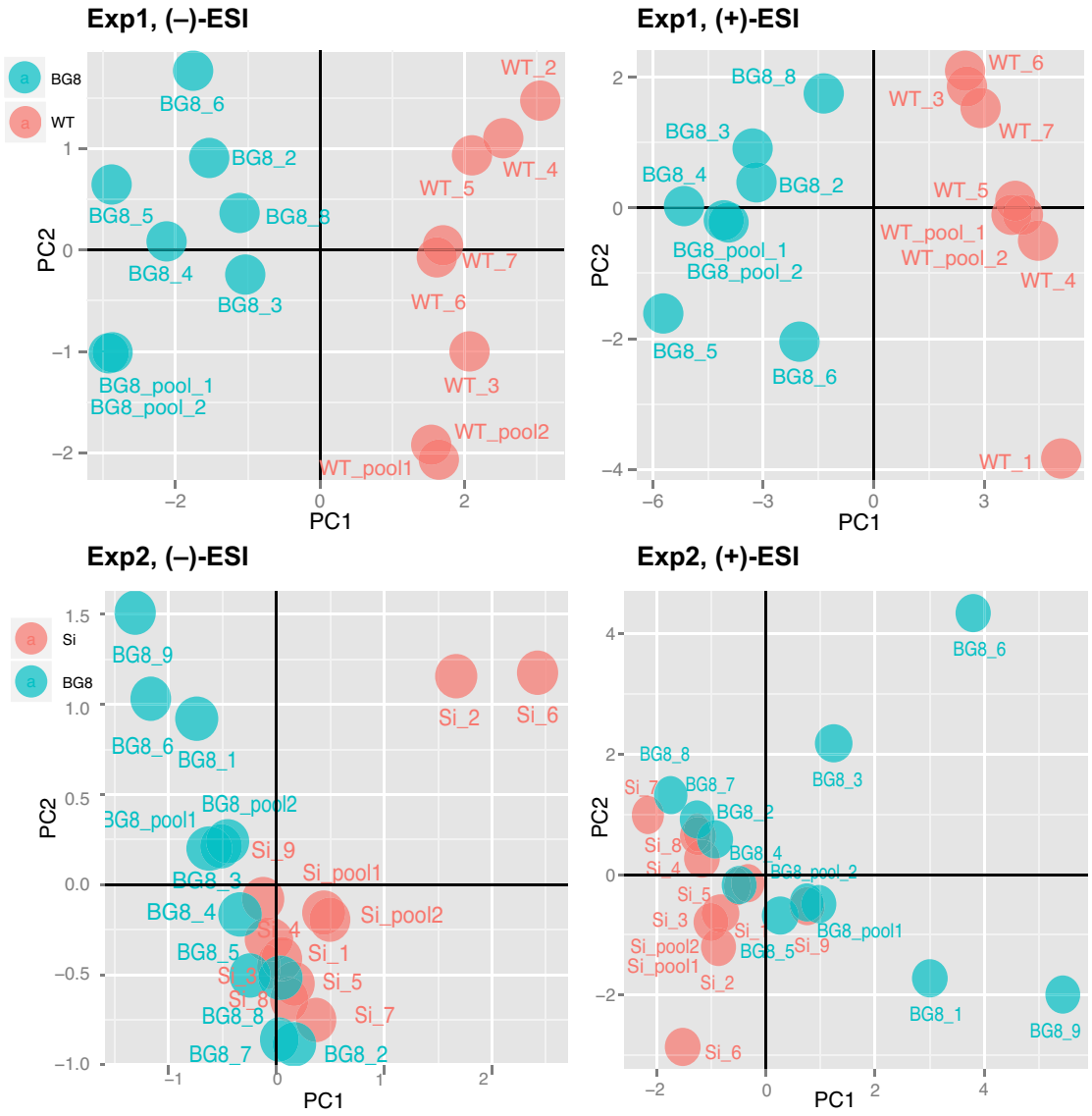
To chapter 2: Metabolomic characterization of the disease resistance gene *Lr34*

Supplementary table 1: Number of features obtained in statistical processing by *XCMS Online* of experiments Exp1-Exp4. Significance was evaluated by Welch's *t*-test for unequal variances and the criterion for significance was *p*-value ≤ 0.05, whereas criterion for fold change was ≥ 1.5.

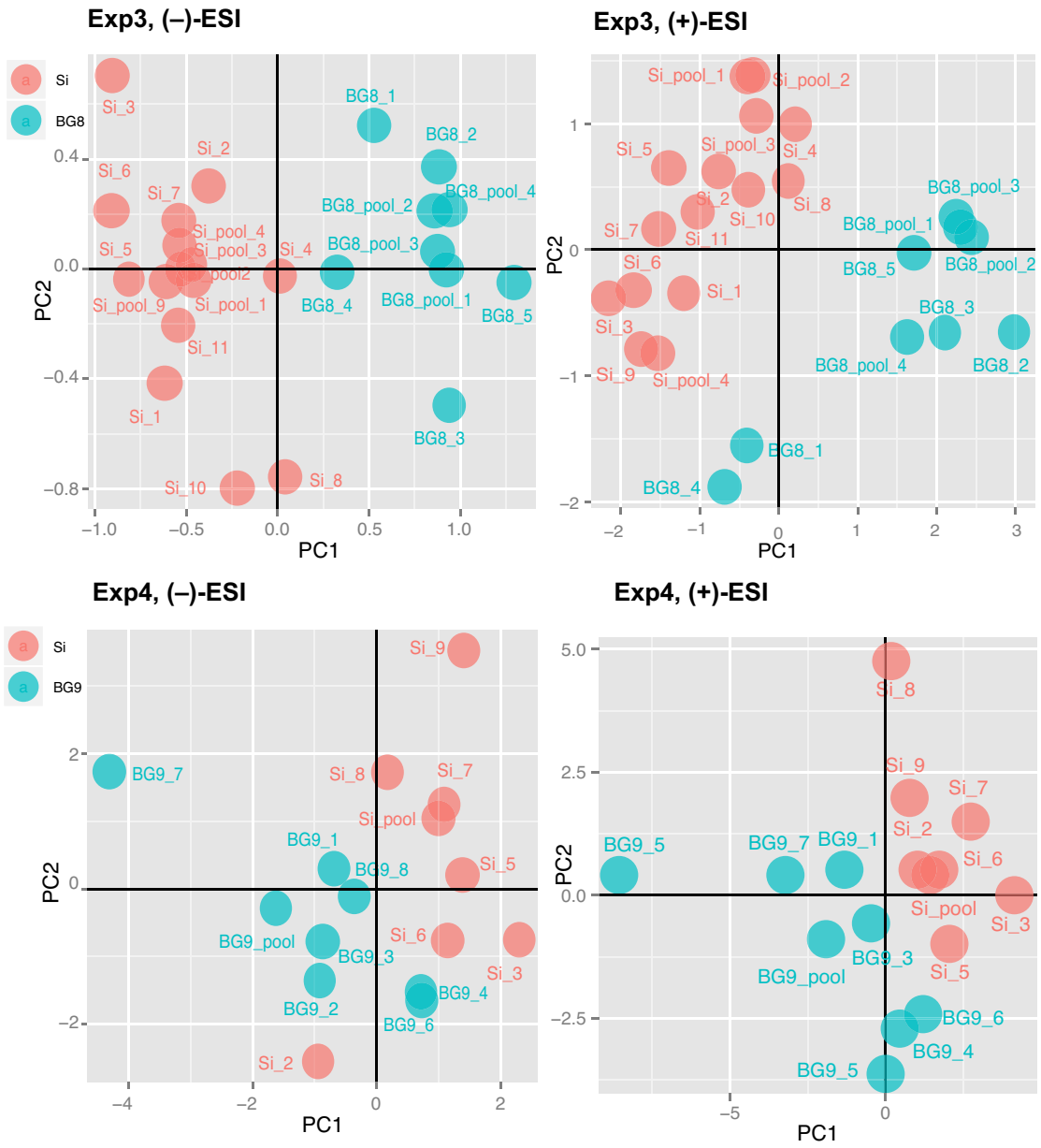
Experiment	ESI	Total	Passed significance criterium	Passed FC criterium
Exp1	(–)	1976	275	195
	(+)	3977	406	312
Exp2	(–)	719	99	34
	(+)	4164	593	133
Exp3	(–)	879	205	121
	(+)	2393	365	240
Exp4	(–)	5185	367	131
	(+)	10655	1061	620

Supplementary table 2: Number of features obtained in statistical processing by *Feature filtering* of experiments Exp1-Exp4. Quality control criterion was CV(samples) < CV(QC) and intensity criterion was intensity > MS/MS threshold (1000). Significance was evaluated by Student's t-test where the criterion was *p*-value ≤ 0.05, and fold change criterion was FC ≥ 1.5.

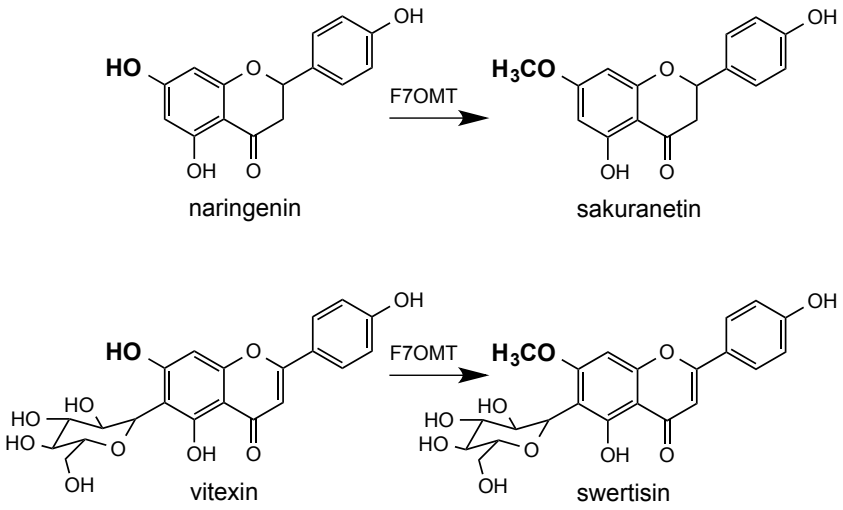
Experiment	ESI	Total	Passed QC + intensity criterium	Passed significance + FC criterium
Exp1	(–)	2723	53	27
	(+)	3273	584	233
Exp2	(–)	297	84	2
	(+)	3278	474	0
Exp3	(–)	233	57	3
	(+)	996	106	32
Exp4	(–)	2143	415	0
	(+)	3548	768	21



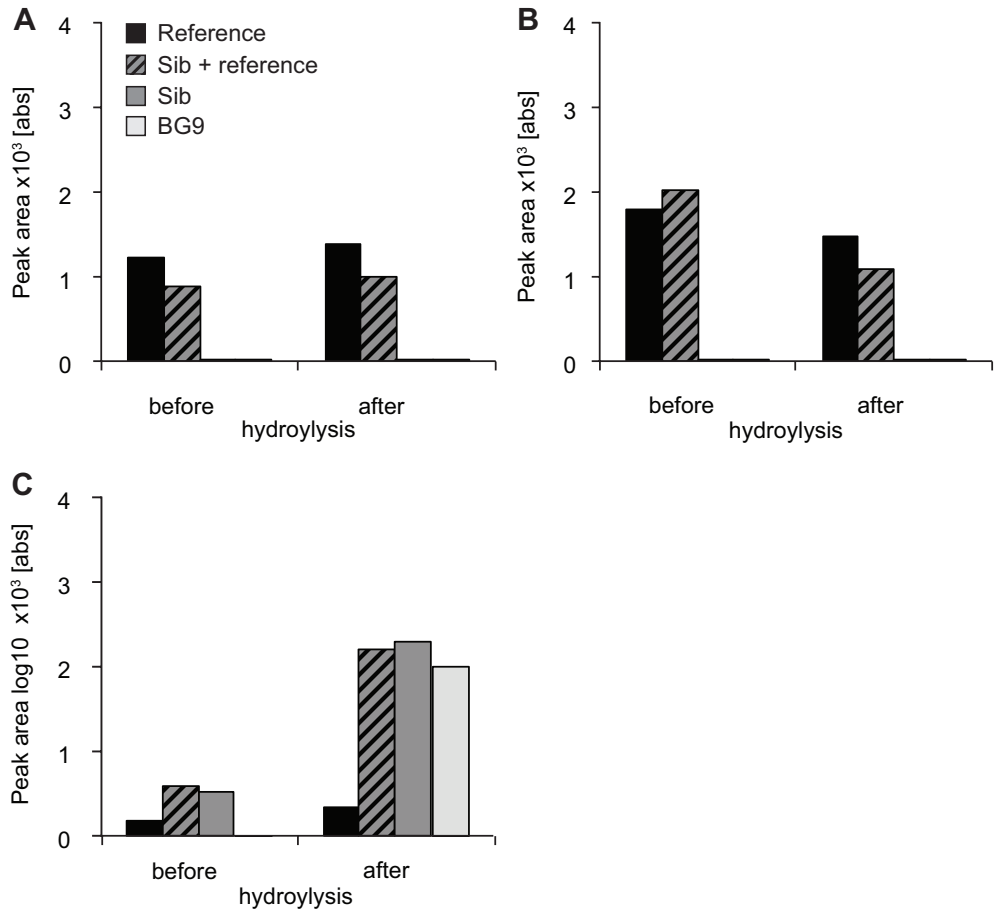
Supplementary figure 1: Principle component analysis (PCA) of Exp1 and Exp2, negative- and positive ionization data. Clear group separation was obtained between control (orange) and transgenic *Lr34* barley (blue) in Exp1 but not in Exp2.



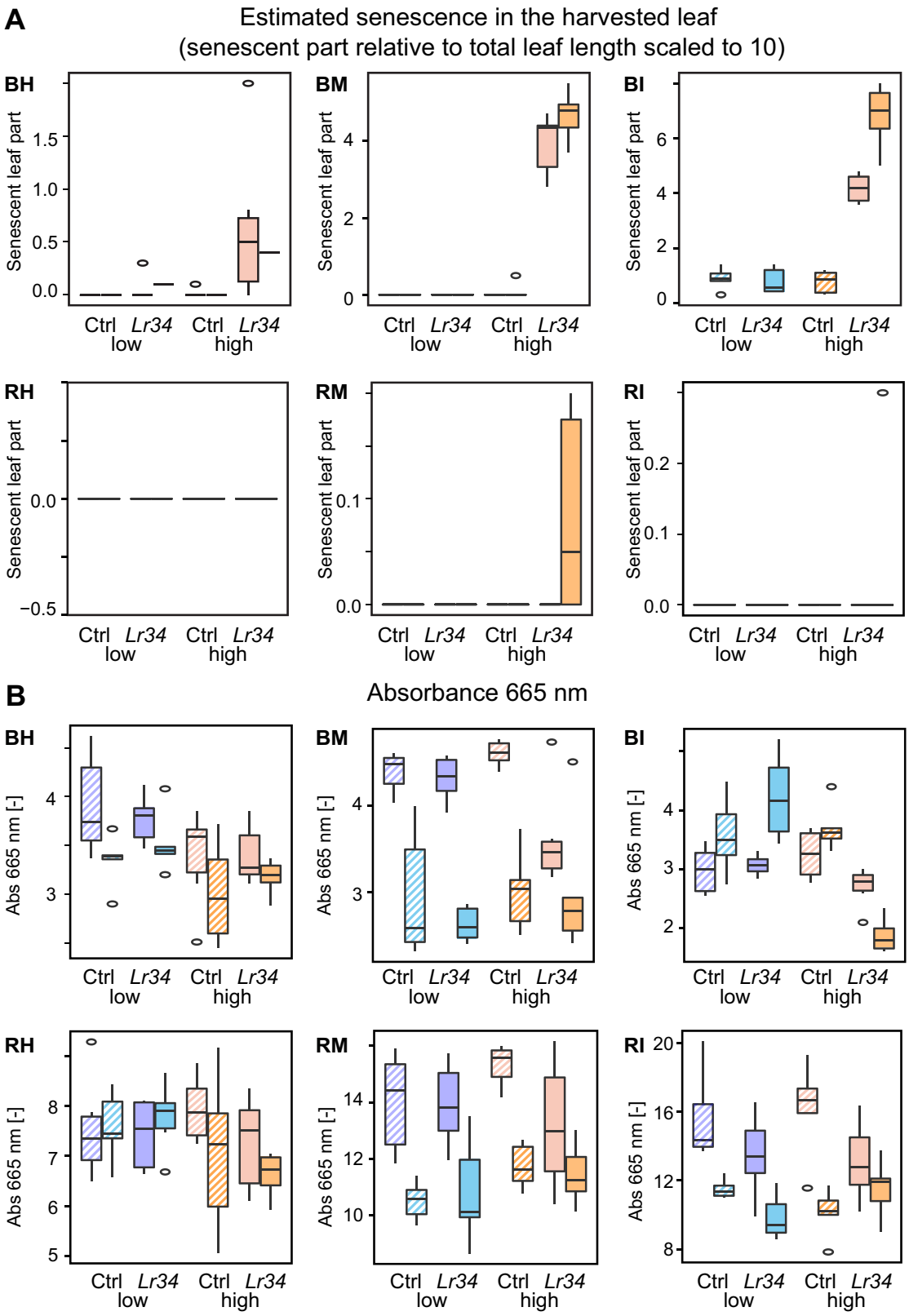
Supplementary figure 2: Principle component analysis (PCA) of Exp3 and Exp4, negative- and positive ionization data. Group separation was obtained between control (orange) and transgenic *Lr34* barley (blue) in Exp3, but not in Exp4.



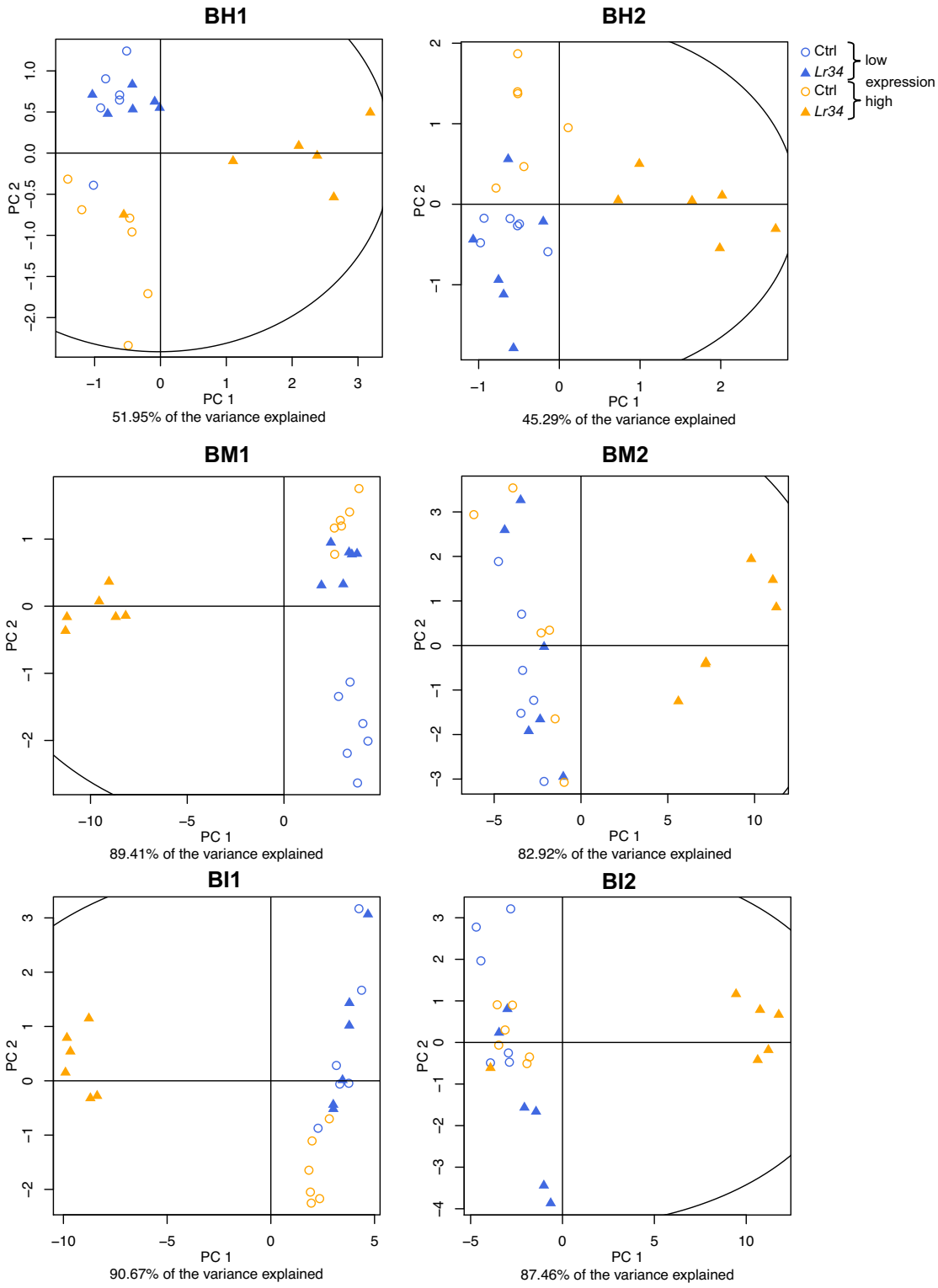
Supplementary scheme 1: Further substrates on F7OMT: naringenin and its methoxylated product sakuranetin, and vitexin (apigenin-6-*C*-glucoside), and its biotransformation product swertisin



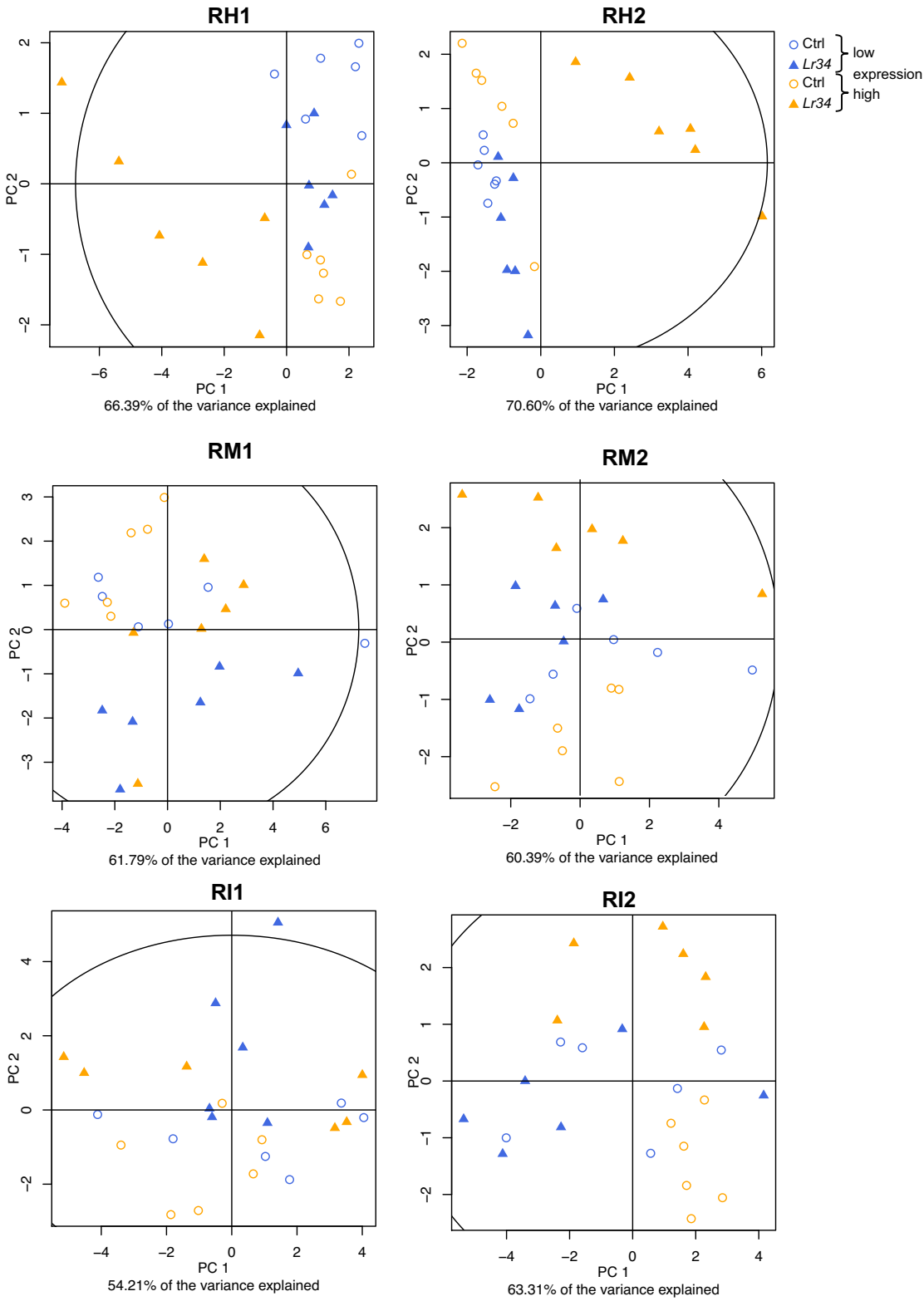
Supplementary figure 3: Hydrolysis of plant material standard addition: reference compounds naringenin (A), its methoxylated product sakuranetin (B), and vitexin (C) with log10-transformed scale. Swertisin was not detected in the extracted ion chromatograms of hydrolyzed plant material.



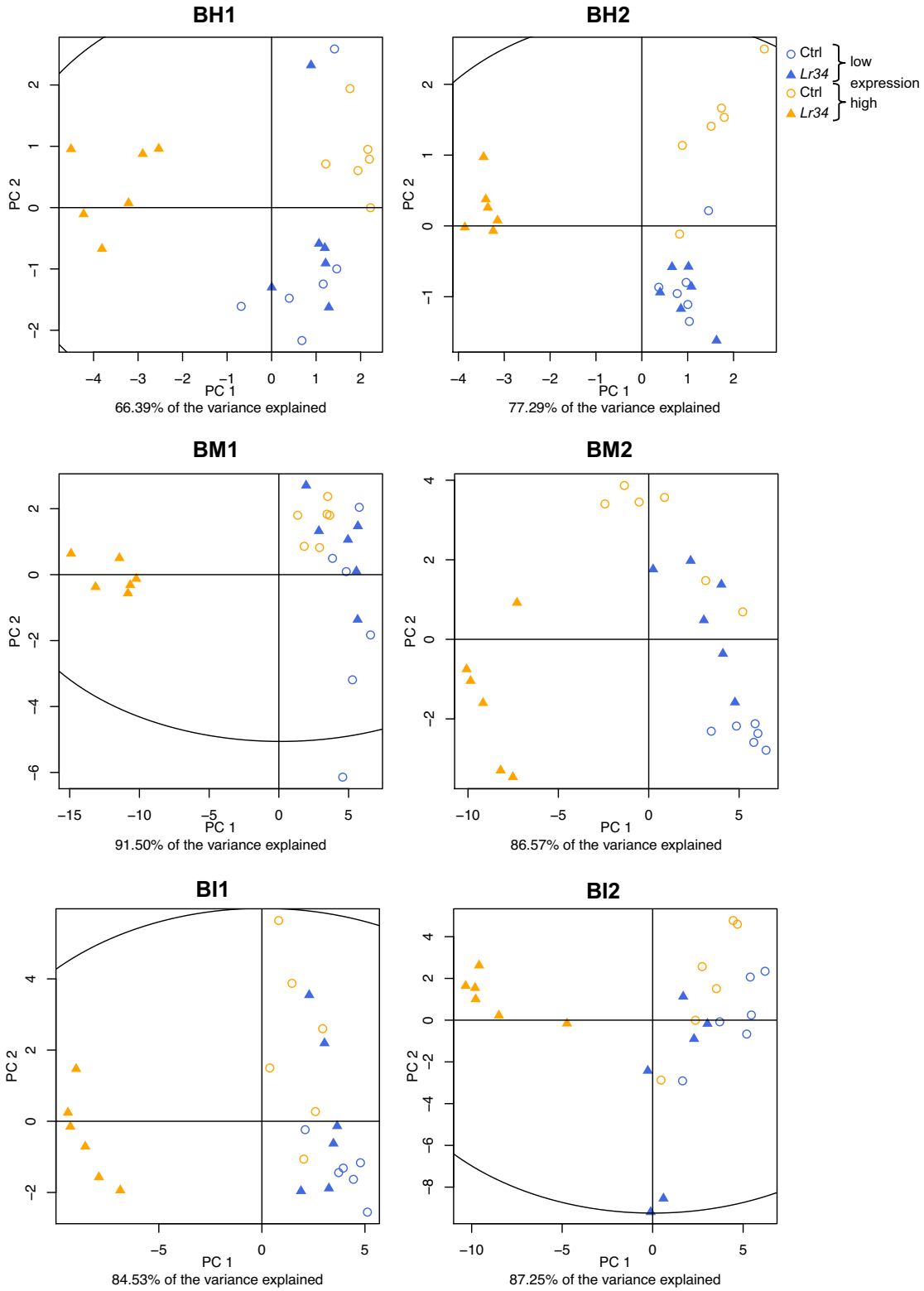
Supplementary figure 4: Evaluation of senescence in barley and rice: estimated senescence based on the senescence part of the harvested leaf relative to the leaf length scaled to 10 (A), and measurement of chlorophyll A based on absorbance at 665 nm (B).



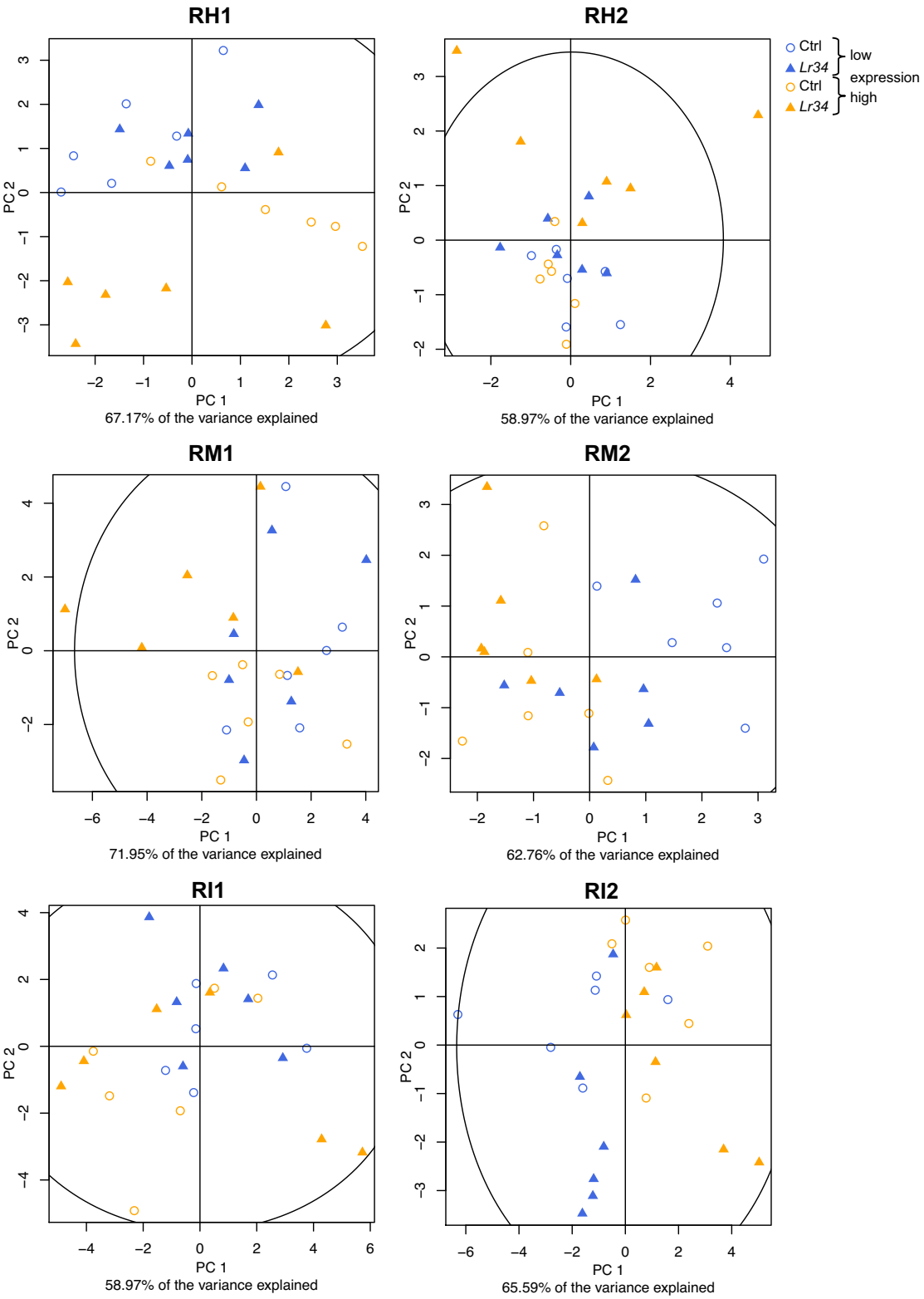
Supplementary figure 5: PCA score plots of the identified primary metabolites in barley. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.



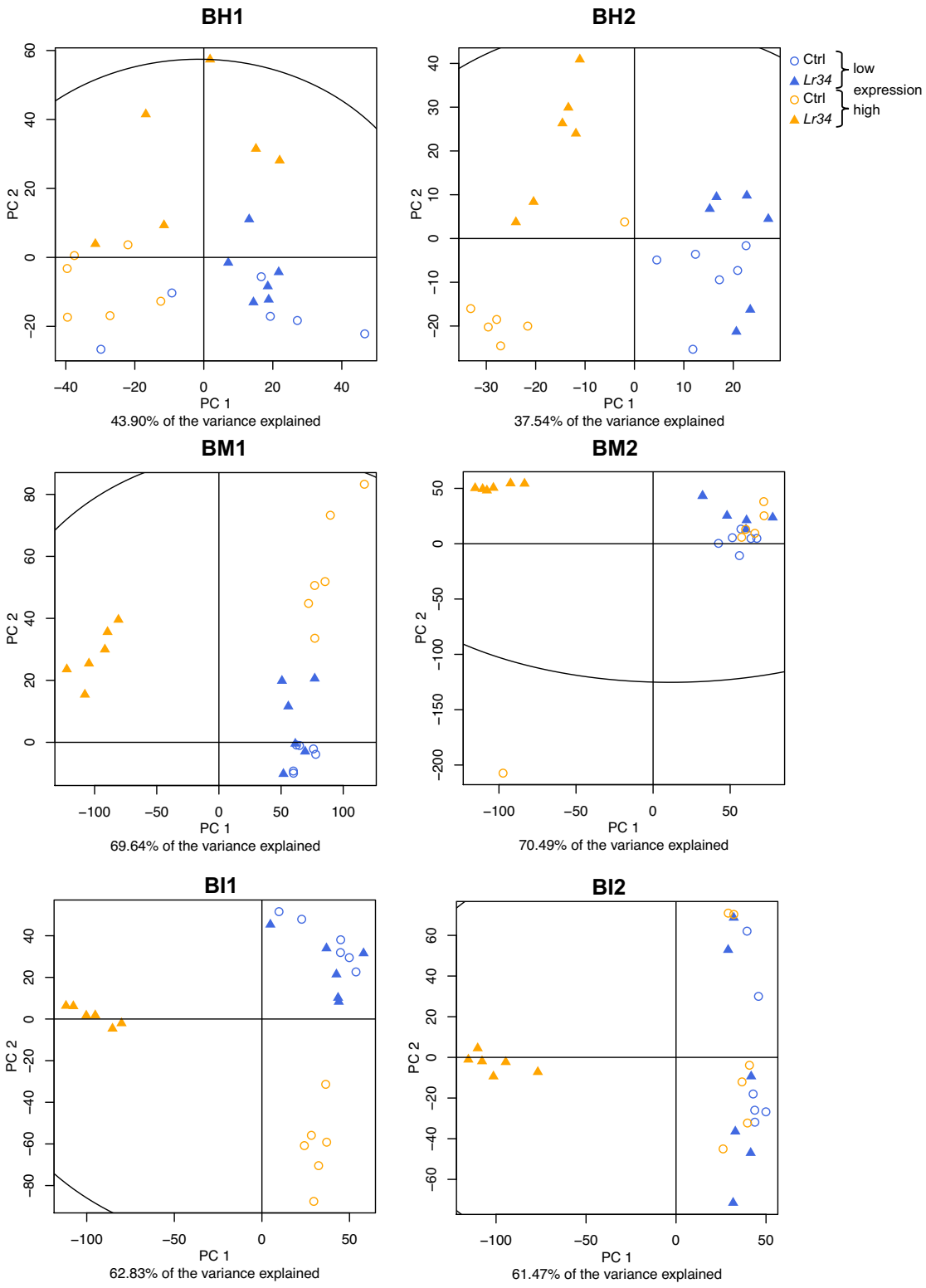
Supplementary figure 6: PCA score plots of the identified primary metabolites in barley. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.



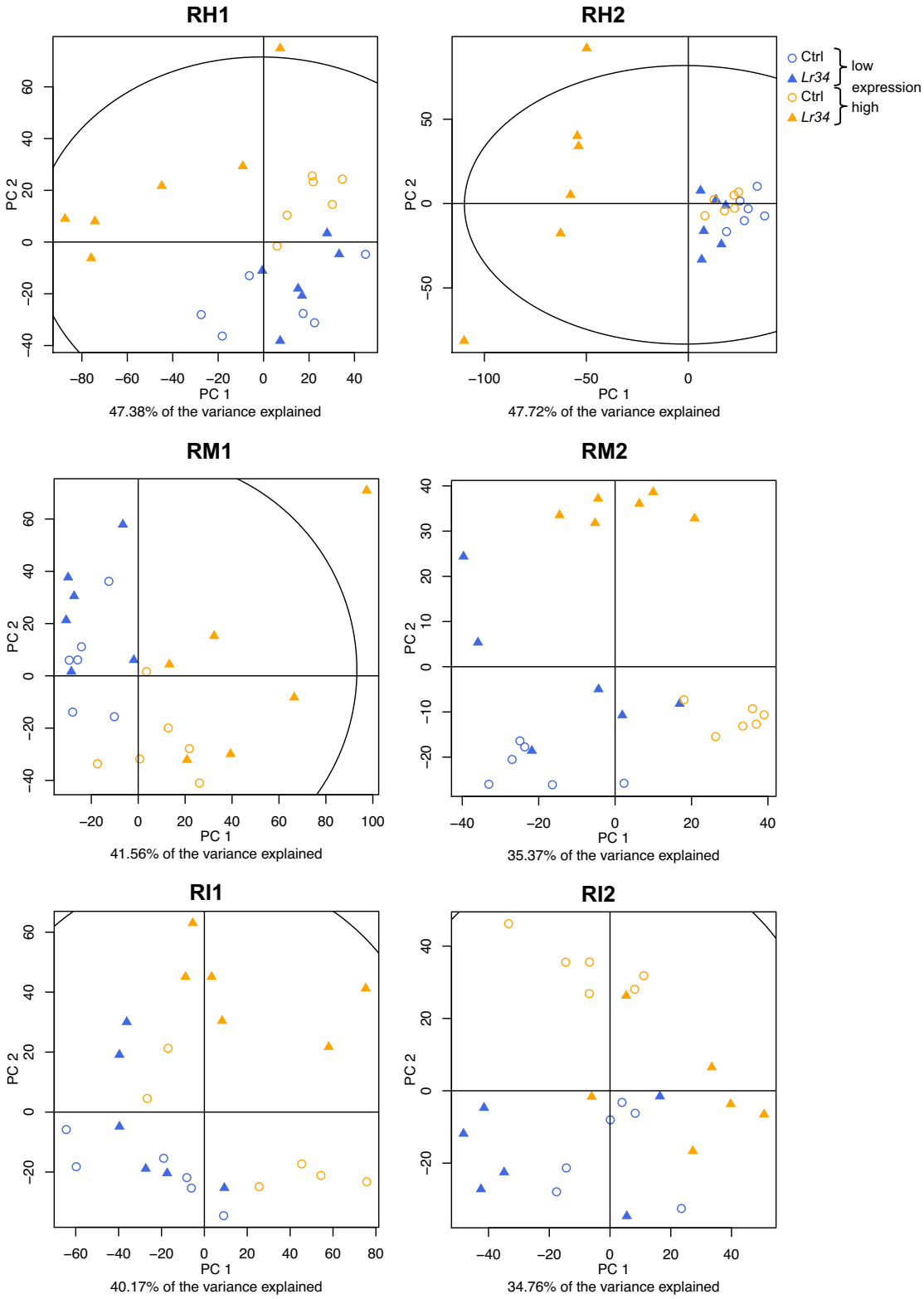
Supplementary figure 7: PCA score plots of the identified lipids in barley. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.



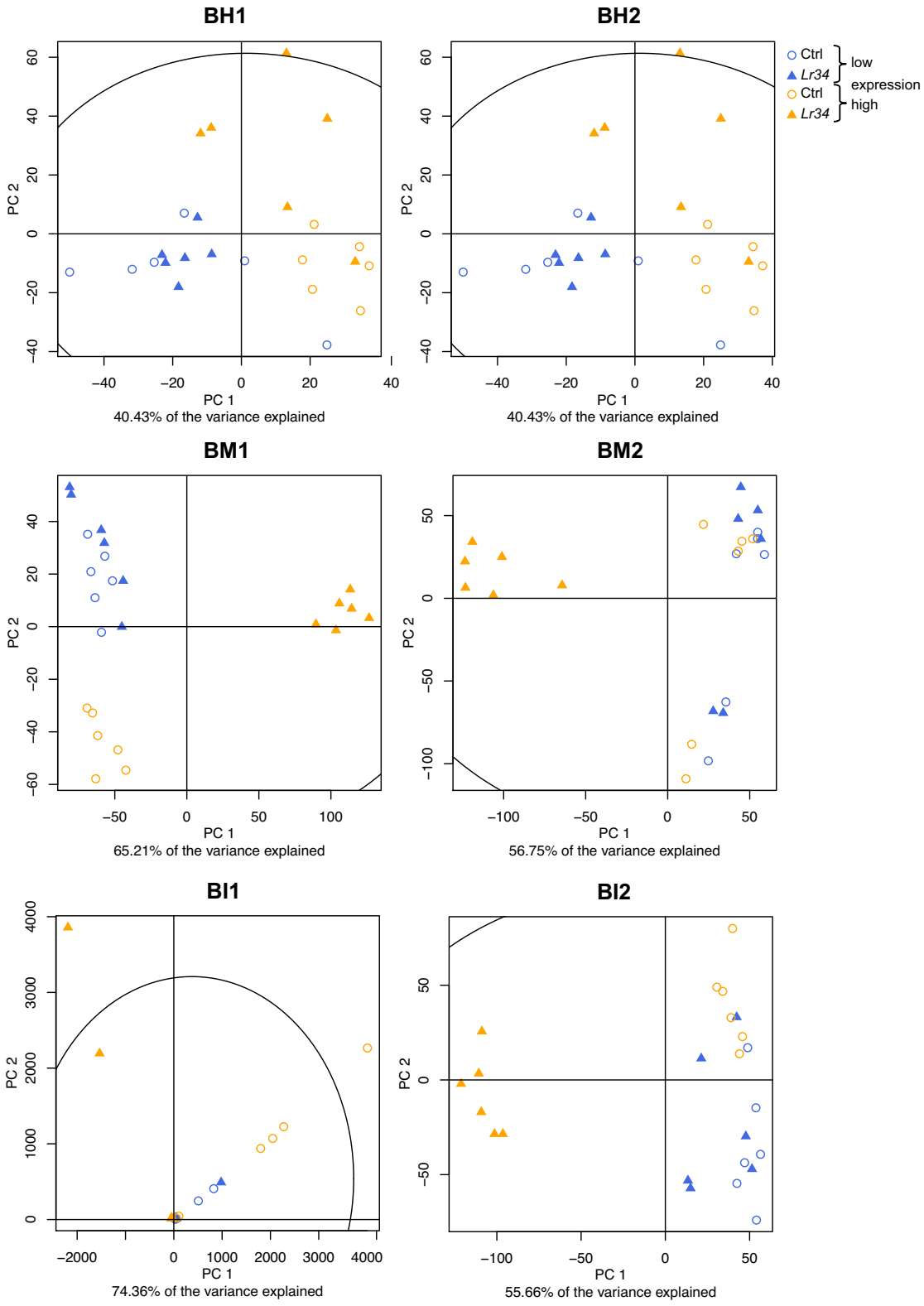
Supplementary figure 8: PCA score plots of the identified lipids in rice. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.



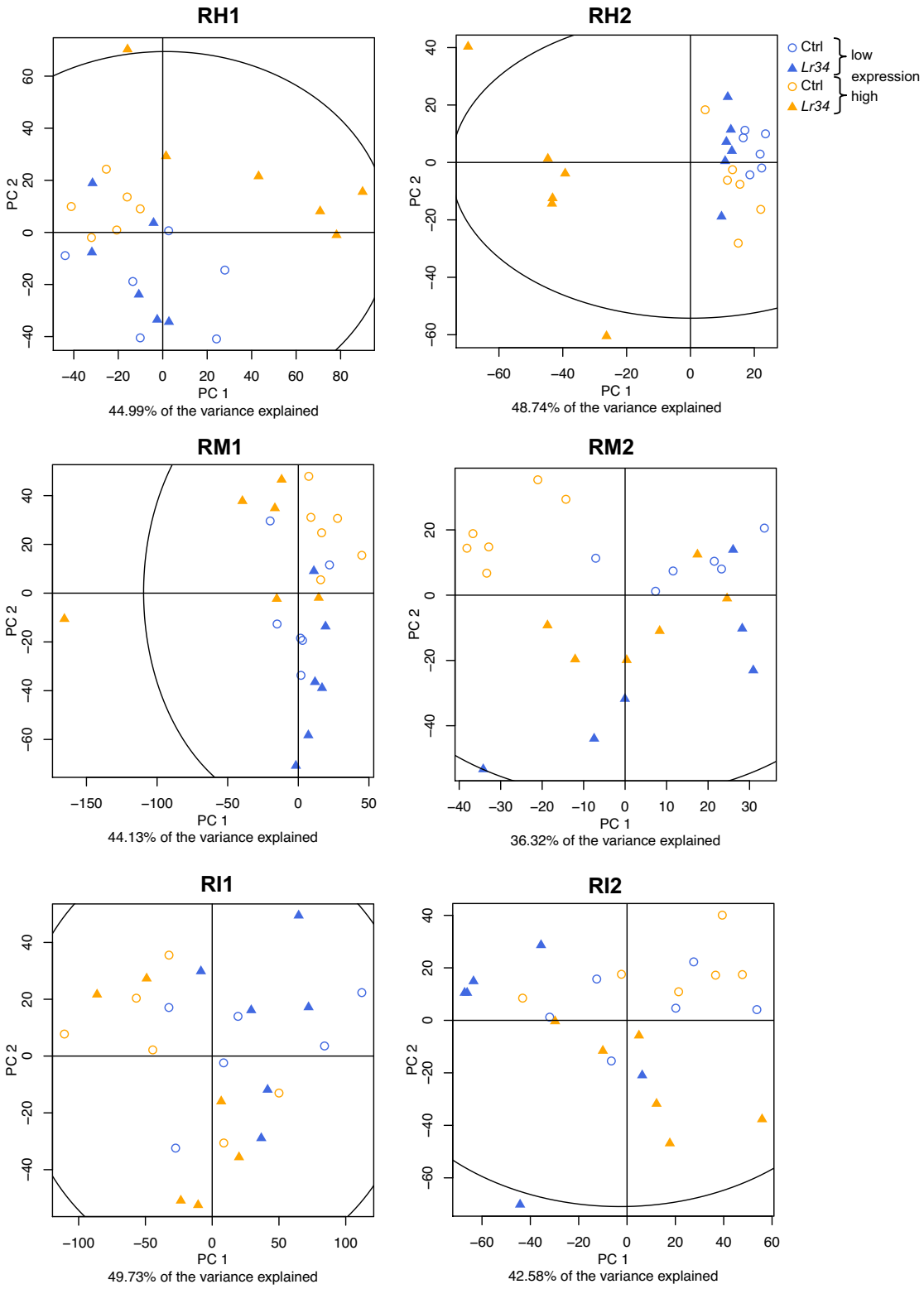
Supplementary figure 9: PCA score plots of the LC-(–)-ESI-MS secondary metabolic features in barley. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.



Supplementary figure 10: PCA score plots of the LC-(–)-ESI-MS secondary metabolic features in rice. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.



Supplementary figure 11: PCA score plots of the LC-(+)-ESI-MS secondary metabolic features in barley. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.

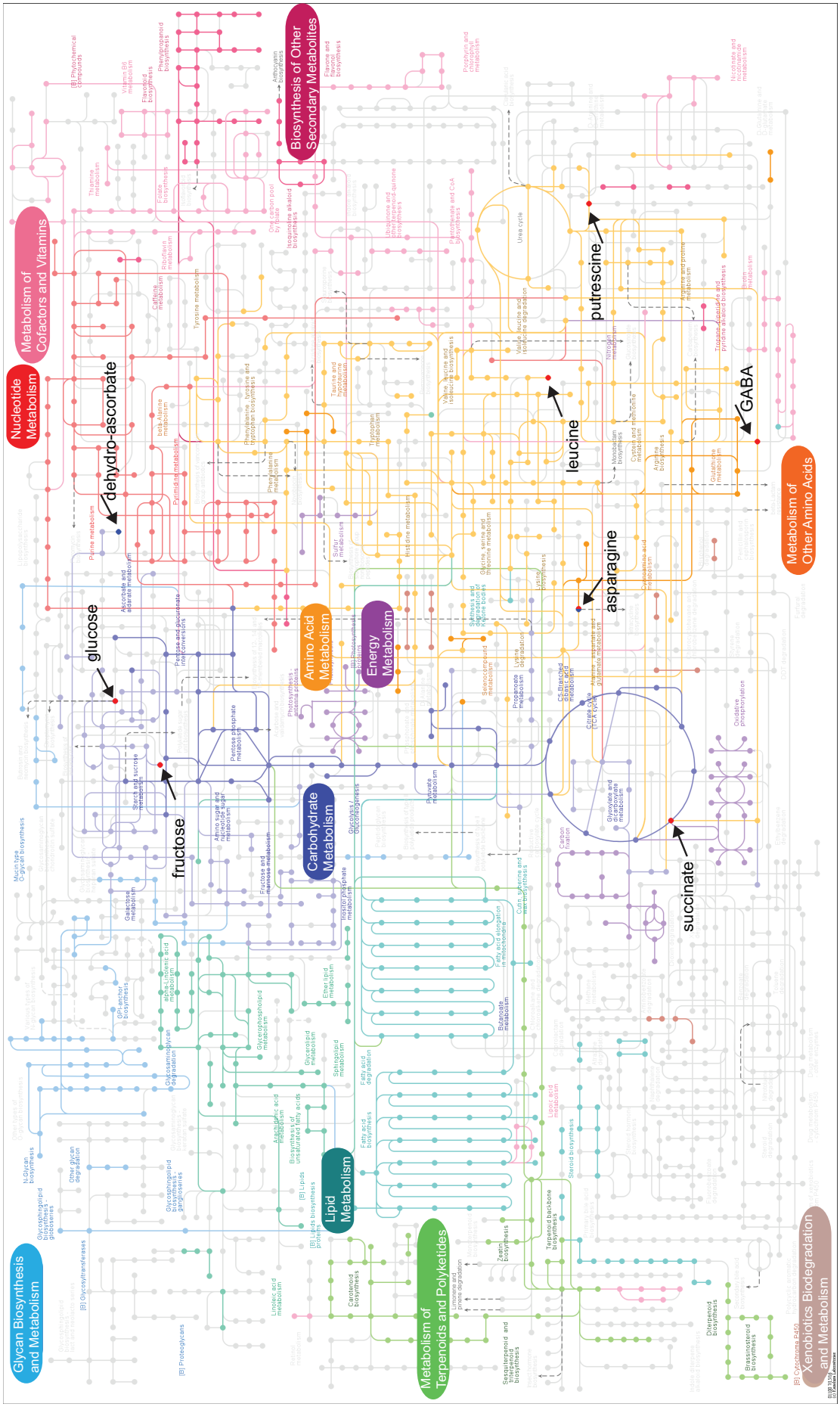


Supplementary figure 12: PCA score plots of the LC-(+)-ESI-MS secondary metabolic features in rice. Low expression levels are symbolized in blue, high expression levels in orange, controls as circles, *Lr34* as triangles.

Supplementary table 3: Primary metabolites detected in barley and rice (B+R) and wheat (W). Metabolites removed in filtering are indicated with (X). Thymine was only partially detected in B+R.

Metabolite	RI	*m/z* Quantification fragment	B+R	W
Amino acids and derivatives				
Alanine	208806	116	X	X
Asparagine	479652	188	X	(X)
Aspartate	458073	232	X	X
Cysteine	479000	220	X	X
Glutamate	508699	246	X	X
Glutamine	598716	156	X	X
Glycine	325397	174	X	X
Histidine	678120	154	X	X
Homoserine	406611	128	X	X
Isoleucine	319474	158	X	X
Leucine	305678	158	X	X
Lysine	615586	156	X	X
Methionine	474411	128	X	X
O-Acetylserine	411982	174		X
Ornitine	570761	142	X	X
Phenylalanine	531003	192	X	X
Proline	338708	142	X	(X)
Pyroglutamate	506770	156	X	X
Serine	358217	204	X	X
Threonine	368104	218	X	X
Thymine	422976	255	(X)	
Tryptophan	790160	202	X	X
Tyramine	638368	174	X	(X)
Tyrosine	658224	218	X	X
Valine	272239	144	X	X
β-Alanine	394629	174	X	X
Sugars and polyols				
1-*O*-Methyl-glucopyranoside	615042	204	X	X
^{13}C-Sorbitol (internal standard)	587733	220	X	X
allo-Inositol	548930	305		X
Arabinose / Lyxose	498474	217	X	X
Cellobiose	859011	361	X	X
Etythritol / Threitol	409750	217	X	X
Fructose / Psicose	580529	103	X	X
Fructose-6-P	766639	315	X	X
Galactinol	939266	191	X	X
Gluconate-6-P	802320	299	X	X
Glucose	591578	343	X	X
Glucose- / Galactose-6-P	779469	299	X	X
Glycerol	292917	117	X	X
Glycerol-3- / -1-P	574363	299	X	X
Gulose / Talose / Altrose	591519	319	X	
Kestose / Raffinose	1033747	361	X	X
Laminarbiose / Maltose	880538	204		X
Maltose	870668	204	X	X
myo-Inositol	654810	318	X	X
myo-Inositol-1-P	789461	318	X	X
Palatinose	909678	361		X
Rhamnose	515875	117		X
Ribitol / Arabitol	502165	205	X	X
Sucrose	843406	103	X	X
Trehalose	879201	361	X	X
Xylose	493772	103	X	X

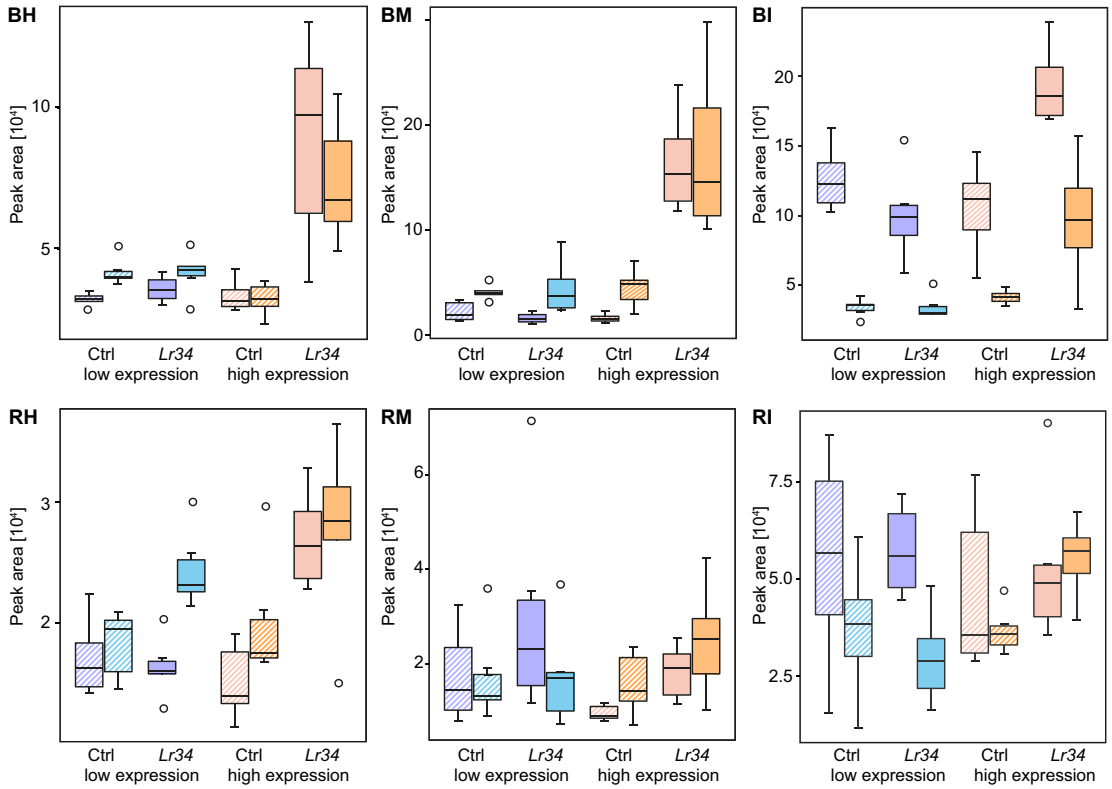
Metabolite	RI	*m/z* Quantification fragment	B+R	W
Organic acids				
1-Dehydro-ascorbate	625416	316	X	X
1,4-Lactone-gluconic acid	646404	217		X
2-Aminoadipate	553570	260	X	X
2-Oxo-glutarate	524062	198	X	X
3-Caffeoyl-quinic acid	1025778	345		X
4-Hydroxy-benzoate	535013	223	X	(X)
γ-Aminobutyric acid (GABA)	452778	174	X	X
Adipate	471285	111	X	
Allantoate	640923	100	X	
Ascorbate	654758	316	X	X
Caffeate	744890	219		X
cis-Aconitate	581596	285	X	X
Citrate	593287	273	X	X
Erythronate	443378	292	X	X
Fumarate	371880	245	X	X
Gluconate / Galactonate	626812	333	X	X
Glycerate	345546	189	X	X
Glycerate-3-P	605665	299	X	X
Isocitrate	597949	375	X	X
Itaconate	386872	259		X
Malate	442318	233	X	X
Maleate	373296	215	X	X
Nicotinate	385967	180	X	X
Oxalate	261778	131	X	
Panthothenate	679363	201		X
Phosphoenolpyruvate	527601	211	X	
Pyruvate	222280	174	X	X
Quinate	578114	345	(X)	X
Salicylate	482578	135	X	
Shikimate	586043	255	X	X
Succinate	365827	247	X	X
Threonate	458816	292	X	X
Fatty acids				
9-(E)-Octadecanoate / Stearate	770993	117	X	(X)
Dodecanoate	530793	117	X	
Hexadecanoate	696677	117	X	X
Nonanoate	378194	117		X
Others				
Adenine	679210	264	X	X
Adenosine-5-P / ADP / ATP	1058965	169	X	
Ethanolamine	284928	174	X	(X)
OrthoP	333868	299	X	(X)
Putrescine	517538	174	X	X
Spermidine	723825	144		X
Sulafate	298583	131	X	
Urea	341248	171	X	X



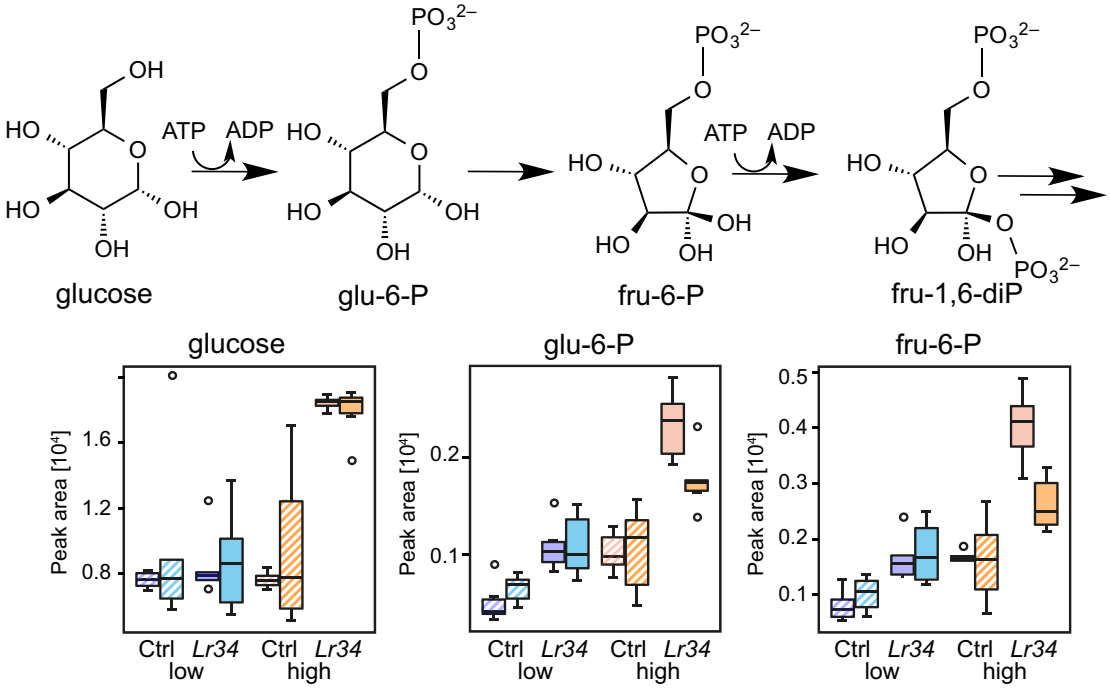
Supplementary figure 13: Mapping of metabolites regulated in high *Lr34*-expressing barley grown under all three growth conditions on KEGG pathway map osa01100 in reference organism *Oryza sativa*; red points are upregulated, blue points downregulated metabolites

Supplementary table 4: Fold change of primary metabolites commonly changed high expression *Lr34* barley mock infected and pathogen infected.

FC experiment	BM1	BM2	BI1	BI2	Mean FC
Cysteine	3.15	2.52	4.78	6.50	4.24
Glutamine	19.39	23.31	16.22	13.62	18.14
Histidine	15.42	12.58	10.33	20.78	14.78
Isoleucine	9.76	8.71	10.34	10.80	9.90
Lysine	5.46	3.43	4.74	4.37	4.50
Methionine	3.30	2.67	4.94	7.80	4.68
Ornithine	6.31	4.30	1.68	2.74	3.75
Phenylalanine	10.27	6.66	7.45	15.72	10.02
Proline	10.49	24.47	12.13	9.16	14.06
Tryptophan	39.61	25.92	14.06	50.23	32.46
Tyrosine	3.39	2.53	1.68	2.47	2.52
Valine	5.59	7.80	10.03	15.48	9.73
β-Alanine	16.08	18.26	12.25	22.37	17.24
2-Aminoadipate	12.25	26.76	3.58	7.54	12.53
Allantoate	1.96	6.79	3.58	17.01	7.34
Erythronate	2.67	1.93	3.01	3.51	2.78
Nicotinate	2.08	1.73	1.74	2.32	1.97
1-O-methyl-glucopyranoside	1.84	2.28	1.58	12.46	4.54
Erythritol / Threitol	4.29	5.41	4.50	16.77	7.75
Fructose-6-phosphate	2.41	1.61	3.40	1.99	2.35
Glucose- / Galactose-6-phosphate	2.29	1.66	3.24	1.95	2.29
Glycerol	5.31	3.01	6.29	11.31	6.48
Glycerol-3-phosphate	2.67	1.72	1.88	2.60	2.22
Gulose / Talose / Altrose	2.85	2.16	4.27	1.92	2.80
myo-Inositol-1-phosphate	2.05	1.87	1.84	2.67	2.11
Threhalose	7.90	4.36	2.34	1.84	4.11
Dodecanoate	19.05	19.50	14.05	14.90	16.88
Adenosin-5-phosphate / ADP / ATP	3.41	4.39	3.10	4.38	3.82
Ethanolamine	1.56	1.58	2.15	2.31	1.90
Sulfate	1.80	1.55	2.03	2.94	2.08
Urea	1.49	1.69	2.26	5.09	2.63



Supplementary figure 14: Succinate (**17**) levels in barley and rice; upregulation in *Lr34* high expression of BH, BM, BI, tendency of increase in RH.



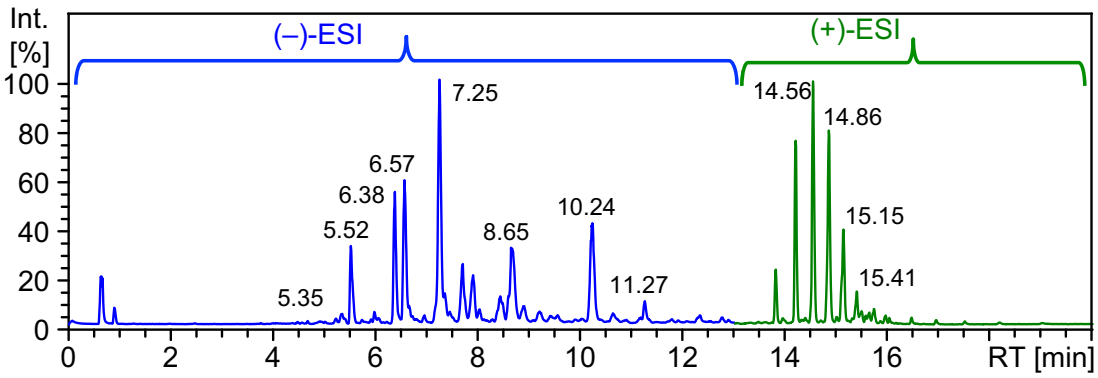
Supplementary figure 15: Primary metabolites involved in glycolysis are upregulated in high- and partially in low-expressing barley plants grown under mock-infected growth condition.

Supplementary table 5: Lipids detected in barley and rice (B+R) and wheat (W).

LipidID	Chemical Formula	ESI	Ion	*m/z* calculated	Verified RT [min]	B+R PeakID	B+R *m/z*	B+R RT [min]	B+R Deviation *m/z*	B+R Deviation RT [min]	W PeakID	W *m/z*	W RT [min]	W Deviation *m/z* [ppm]	W Deviation RT [min]
Cer t18:0/c22:0	C40H81NO4	(-)	[M+CH3COO]-	698.63041	12.08	19529	698.63176	11.74	-1.94	0.34	14713	698.63180	11.74	-1.99	0.33
Cer t18:0/h16:0	C34H69NO5	(-)	[M+CH3COO]-	630.53143	8.11	16360	630.53165	8.06	-0.36	0.05	12620	630.53255	8.05	-1.79	0.07
Cer t18:0/h24:1	C42H83NO5	(-)	[M+CH3COO]-	740.64098	11.70	-	NA	NA	NA	NA	16093	740.64257	11.90	-2.16	-0.20
Cer t18:1/c26:0	C44H87NO4	(+)	[M + H]+	694.67079	14.11	6969	694.67749	13.98	-9.64	0.13	-	NA	NA	NA	NA
Cer t18:1/h22:0	C40H79NO5	(-)	[M+CH3COO]-	712.60968	11.10	-	NA	NA	NA	NA	15129	712.61128	10.84	-2.25	0.27
Cer t18:1/h24:0	C42H83NO5	(+)	[M + H]+	682.63440	12.42	6701	682.63666	12.33	-3.31	0.08	7618	682.63698	12.30	-3.77	0.11
DGDG 32:0	C47H88O15	(-)	[M+CH3COO]-	951.62673	8.91	31776	951.62881	8.91	-2.18	-0.01	14349	951.62819	8.92	-1.53	-0.01
DGDG 32:1	C47H86O15	(-)	[M+CH3COO]-	949.61108	8.05	31687	949.61278	8.15	-1.79	-0.10	23927	949.61069	8.15	0.41	-0.09
DGDG 32:2	C47H84O15	(-)	[M+CH3COO]-	947.59543	7.38	31571	947.59077	7.40	4.91	-0.03	-	NA	NA	NA	NA
DGDG 32:3	C47H82O15	(-)	[M+CH3COO]-	945.57978	6.74	31495	945.58136	6.76	-1.67	-0.02	23805	945.58074	6.76	-1.02	-0.02
DGDG 34:1	C49H90O15	(-)	[M+CH3COO]-	977.64238	9.12	32794	977.64342	9.14	-1.07	-0.02	14951	977.64320	9.14	-0.84	-0.02
DGDG 34:2	C49H88O15	(-)	[M+CH3COO]-	975.62673	8.33	32714	975.62689	8.37	-0.16	-0.04	14903	975.62805	8.36	-1.35	-0.03
DGDG 34:3	C49H86O15	(-)	[M+CH3COO]-	973.61108	7.67	32651	973.61015	7.68	0.95	-0.01	14862	973.61222	7.68	-1.17	-0.01
DGDG 34:4	C49H84O15	(-)	[M+CH3COO]-	971.59543	6.94	32592	971.59715	6.95	-1.77	-0.01	24615	971.59671	6.96	-1.31	-0.02
DGDG 34:5	C49H82O15	(-)	[M+CH3COO]-	969.57978	6.39	32530	969.57987	6.43	-0.09	-0.04	-	NA	NA	NA	NA
DGDG 34:6	C49H80O15	(-)	[M+CH3COO]-	967.56413	5.81	32441	967.56624	5.85	-2.18	-0.03	24488	967.56612	5.84	-2.05	-0.03
DGDG 36:1	C51H94O15	(-)	[M+CH3COO]-	1005.67368	10.16	33764	1005.67500	10.16	-1.31	0.00	-	NA	NA	NA	NA
DGDG 36:2	C51H92O15	(-)	[M+CH3COO]-	1003.65803	9.39	33708	1003.65940	9.39	-1.36	0.00	15596	1003.65942	9.40	-1.38	-0.01
DGDG 36:3	C51H90O15	(-)	[M+CH3COO]-	1001.64238	8.53	33642	1001.64007	8.55	2.31	-0.02	25484	1001.64152	8.70	0.86	-0.17
DGDG 36:4	C51H88O15	(-)	[M+CH3COO]-	999.62673	7.82	33563	999.62761	7.82	-0.88	0.00	15518	999.62827	7.83	-1.53	-0.01
DGDG 36:5	C51H86O15	(-)	[M+CH3COO]-	997.61108	7.15	33484	997.61188	7.16	-0.80	-0.01	15470	997.61187	7.16	-0.79	-0.01
DGDG 36:6	C51H84O15	(-)	[M+CH3COO]-	995.59543	6.42	33415	995.59455	6.56	0.89	-0.15	15429	995.59517	6.57	0.26	-0.15
DGDG 38:4	C53H92O15	(-)	[M+CH3COO]-	1027.65803	8.87	34377	1027.65906	8.84	-1.00	0.03	26056	1027.66003	8.85	-1.94	0.02
DGDG 38:5	C53H90O15	(-)	[M+CH3COO]-	1025.64238	8.02	34310	1025.63969	8.03	2.62	-0.02	26002	1025.63942	8.03	2.88	-0.02
DGDG 38:6	C53H88O15	(-)	[M+CH3COO]-	1023.62673	7.24	34258	1023.62729	7.37	-0.54	-0.13	25950	1023.62837	7.37	-1.60	-0.13
GlcCer t18:1/h16:0	C40H77NO10	(-)	[M+CH3COO]-	790.56860	7.40	-	NA	NA	NA	NA	17871	790.57022	7.02	-2.05	0.38
GlcCer t18:1/h22:0	C46H89NO10	(-)	[M+CH3COO]-	874.66250	10.63	-	NA	NA	NA	NA	21522	874.66468	10.15	-2.50	0.47
GlcCer t18:1/h24:0	C48H93NO10	(-)	[M+CH3COO]-	902.69380	11.90	-	NA	NA	NA	NA	22497	902.69483	11.25	-1.14	0.65
GlcCer d or t18:1/c24:1	C48H91NO9	(-)	[M+CH3COO]-	884.68324	11.56	-	NA	NA	NA	NA	21858	884.68572	11.69	-2.81	-0.13
MGDG 32:0	C41H78O10	(-)	[M+CH3COO]-	789.57390	9.77	23900	789.57438	9.77	-0.61	0.00	17839	789.57380	9.78	0.13	-0.02
MGDG 32:1 (1)	C41H76O10	(-)	[M+CH3COO]-	787.55825	8.97	23793	787.55773	8.97	0.67	0.00	-	NA	NA	NA	NA
MGDG 32:2 (2)	C41H74O10	(-)	[M+CH3COO]-	785.54260	8.15	23670	785.54006	8.16	3.23	0.00	17691	785.54174	8.16	1.10	-0.01
MGDG 32:3	C41H72O10	(-)	[M+CH3COO]-	783.52695	7.47	23532	783.52606	7.49	1.14	-0.02	17596	783.52338	7.50	4.56	-0.03
MGDG 34:1	C43H80O10	(-)	[M+CH3COO]-	815.58955	9.96	25399	815.59089	9.98	-1.64	-0.02	18999	815.59042	9.96	-1.07	0.00
MGDG 34:2	C43H78O10	(-)	[M+CH3COO]-	813.57390	9.16	25296	813.57493	9.18	-1.26	-0.02	-	NA	NA	NA	NA
MGDG 34:3	C43H76O10	(-)	[M+CH3COO]-	811.55825	8.47	25172	811.55898	8.47	-0.89	0.00	10179	811.55949	8.47	-1.53	-0.01
MGDG 34:4	C43H74O10	(-)	[M+CH3COO]-	809.54260	7.67	25046	809.54295	7.68	-0.43	-0.01	10126	809.54413	7.68	-1.88	-0.02
MGDG 34:5	C43H72O10	(-)	[M+CH3COO]-	807.52695	7.09	24916	807.51359	7.02	16.54	0.07	-	NA	NA	NA	NA
MGDG 34:6	C43H70O10	(-)	[M+CH3COO]-	805.51130	6.49	24794	805.51227	6.48	-1.20	0.01	18536	805.51260	6.48	-1.62	0.01
MGDG 36:2	C45H82O10	(-)	[M+CH3COO]-	841.60520	10.16	26810	841.60438	10.21	0.98	-0.05	-	NA	NA	NA	NA
MGDG 36:3	C45H80O10	(-)	[M+CH3COO]-	839.58955	9.37	26692	839.59134	9.37	-2.13	-0.01	-	NA	NA	NA	NA
MGDG 36:4	C45H78O10	(-)	[M+CH3COO]-	837.57390	8.60	26582	837.57357	8.59	0.39	0.01	11001	837.57437	8.60	-0.56	-0.01
MGDG 36:5 (1)	C45H76O10	(-)	[M+CH3COO]-	835.55825	7.89	26481	835.55838	7.90	-0.15	0.00	10935	835.55887	7.90	-0.74	-0.01
MGDG 36:5 (2)	C45H76O10	(-)	[M+CH3COO]-	835.55825	7.98	26482	835.55919	8.05	-1.12	-0.06	-	NA	NA	NA	NA
MGDG 36:6	C45H74O10	(-)	[M+CH3COO]-	833.54260	7.29	263881	833.54262	7.25	-0.02	0.04	10883	833.54290	7.25	-0.35	0.04
MGDG 38:4	C47H82O10	(-)	[M+CH3COO]-	865.60520	9.64	28060	865.60702	9.66	-2.10	-0.02	-	NA	NA	NA	NA
MGDG 38:6	C47H78O10	(-)	[M+CH3COO]-	861.57390	8.11	27847	861.57515	8.11	-1.44	0.00	11757	861.57527	8.12	-1.58	-0.01
PG 32:1	C38H73O10P	(-)	[M-H]-	719.48686	7.67	20422	719.48742	7.41	-0.78	0.26	15404	719.48835	7.41	-2.07	0.26
PG 34:2	C40H75O10P	(-)	[M-H]-	745.50251	7.94	-	NA	NA	NA	NA	16273	745.50398	7.60	-1.98	0.34
PG 34:4	C40H71O10P	(-)	[M-H]-	741.47121	7.18	21442	741.47218	6.37	-1.31	0.81	16108	741.47198	6.35	-1.04	0.83
PI 34:3	C43H77O13P	(-)	[M-H]-	831.50290	6.18	-	NA	NA	NA	NA	19751	831.50267	6.36	0.28	-0.18
PI 36:6	C45H75O13P	(-)	[M-H]-	853.48725	5.18	-	NA	NA	NA	NA	20684	853.47968	4.55	8.88	0.63
PC 32:0	C40H80NO8P	(-)	[M+CH3COO]-	792.57656	8.97	24062	792.57807	8.99	-1.91	-0.02	9600	792.57826	8.99	-2.15	-0.02
PC 32:1	C40H78NO8P	(-)	[M+CH3COO]-	790.56103	8.21	23948	790.56175	8.23	-0.92	-0.02	9555	790.56238	8.22	-1.72	-0.01
PC 32:2	C40H76NO8P	(-)	[M+CH3COO]-	788.54538	7.45	23836	788.54626	7.45	-1.12	0.00	9518	788.54727	7.45	-2.40	-0.01
PC 32:3	C40H74NO8P	(-)	[M+CH3COO]-	786.52973	6.80	23715	786.53025	6.80	-0.67	0.00	17720	786.53166	6.81	-2.46	-0.01
PC 34:1	C42H82NO8P	(-)	[M+CH3COO]-	818.59233	9.24	255911	818.59412	9.23	-2.20	0.01	10414	818.59396	9.24	-2.00	-0.01
PC 34:2	C42H80NO8P	(-)	[M+CH3COO]-	816.57668	8.43	254521	816.57663	8.42	0.06	0.01	10349	816.57673	8.44	-0.06	-0.01
PC 34:4	C42H76NO8P	(-)	[M+CH3COO]-	812.54537	6.99	25230	812.54566	7.12	-0.36	-0.13	18824	812.54708	7.01	-2.11	-0.02
PC 36:1	C44H86NO8P	(-)	[M+CH3COO]-	846.62363	10.13	-	NA	NA	NA	NA	11313	846.62432	10.33	-0.82	-0.20
PC 34:6	C42H72NO8P	(-)	[M+CH3COO]-	808.51408	5.91	249671	808.51434	5.92	-0.33	0.00	-	NA	NA	NA	NA
PC 36:2	C44H84NO8P	(-)	[M+CH3COO]-	844.60798	8.96	-	NA	NA	NA	NA	-	NA	NA	NA	NA
PC 36:3	C44H82NO8P	(-)	[M+CH3COO]-	842.59233	8.60	268631	842.59231	8.62	0.01	-0.02	11172	842.59280	8.63	-0.56	-0.03
PC 36:4	C44H80NO8P	(-)	[M+CH3COO]-	840.57668	7.67	26748	840.57612	7.88	0.66	-0.21	-	NA	NA	NA	NA
PC 36:5	C44H78NO8P	(-)	[M+CH3COO]-	838.56103	7.20	266271	838.55895	7.19	2.48	0.01	11032	838.56025	7.20	0.92	0.00
PC 36:6	C44H76NO8P	(-)	[M+CH3COO]-	836.54538	6.43	26519	836.54586	6.61	-0.58	-0.18	10962	836.54561	6.61	-0.28	-0.18
PC 38:2	C46H88NO8P	(-)	[M+CH3COO]-	872.63928	10.13	28429	872.64101	10.12	-1.99	0.01	-	NA	NA	NA	NA
PC 38:3 (1)	C46H86NO8P	(-)	[M+CH3COO]-	870.62363	9.39	283031	870.62492	9.38	-1.49	0.01	-	NA	NA	NA	NA
PC 38:3 (2)	C46H86NO8P	(-)	[M+CH3COO]-	870.62363	9.61	283044	870.62251	9.59	1.28	0.02	-	NA	NA	NA	NA
lysoPC 18:1a	C26H52NO7P	(-)	[M+CH3COO]-	580.36199	4.32	14426	580.36240	3.11	-0.70	1.21	11365	580.36269	3.11	-1.20	1.20
lysoPC 18:1b	C26H52NO7P	(-)	[M+CH3COO]-	580.36199	4.47	-	NA	NA	NA	NA	11366	580.36203	3.28	-0.07	1.20
lysoPC 18:2a	C26H50NO7P	(-)	[M+CH3COO]-	578.34634	3.99	-	NA	NA	NA	NA	11311	578.34742	2.54	-1.86	1.45
lysoPC 18:2b	C26H50NO7P	(-)	[M+CH3COO]-	578.34634	3.82	14370	578.34707	2.69	-1.25	1.13	11312	578.34740	2.70	-1.82	1.12
lysoPC 18:3a	C26H48NO7P	(-)	[M+CH3COO]-	576.33069	3.52	14287	576.33122	2.03	-0.91	1.50	11251	576.33178	2.03	-1.88	1.49
lysoPC 18:3b	C26H48NO7P	(-)	[M+CH3COO]-	576.33069	3.34	14288	576.33147	2.18	-1.35	1.16	11252	576.33164	2.19	-1.63	1.15

Supplementary table 5, *continued*

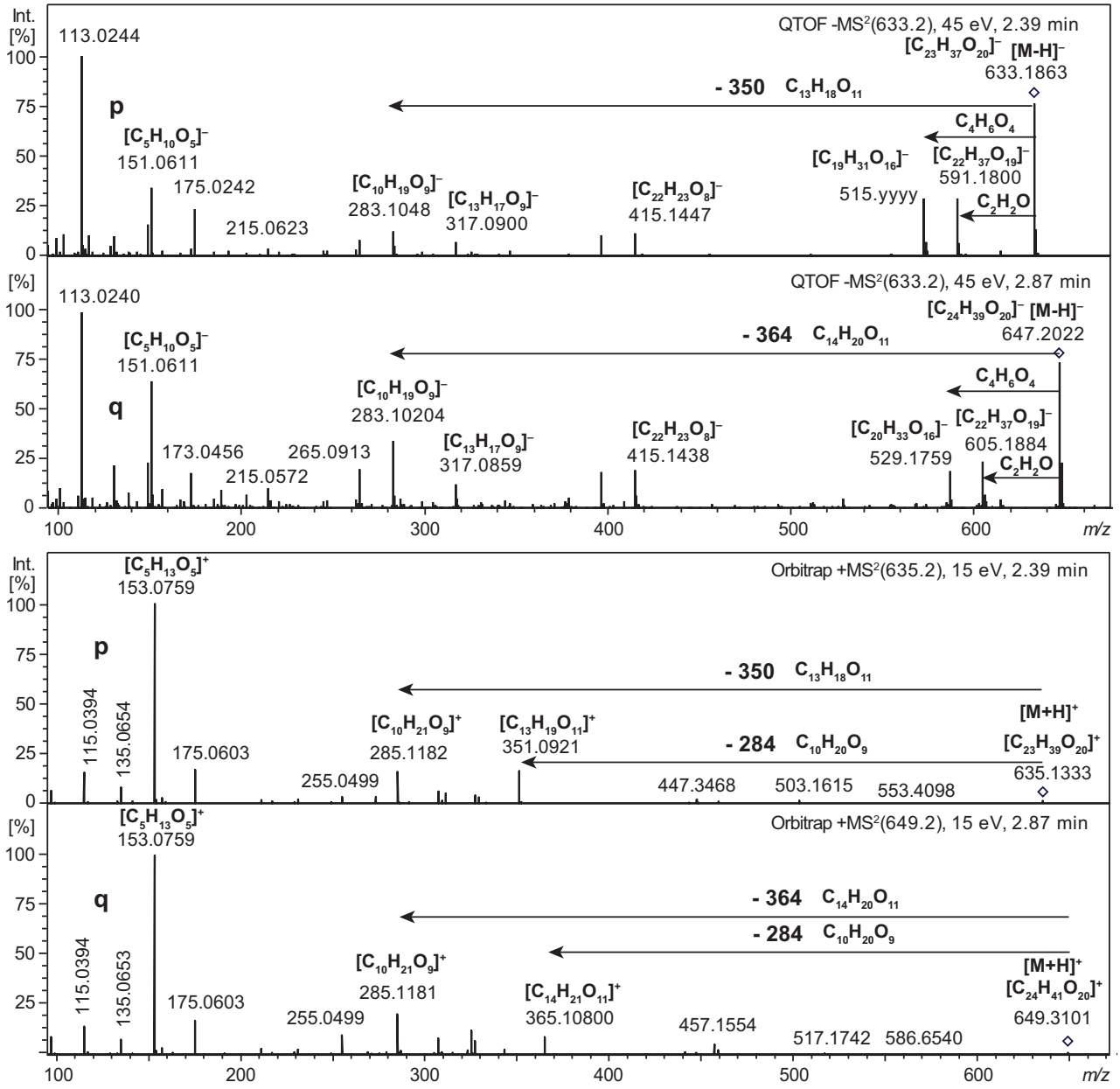
LipidID	Chemical Formula	ESI	Ion	*m/z* calculated	Verified RT [min]	B+R PeakID	B+R *m/z*	B+R RT [min]	B+R Deviation *m/z*	B+R Deviation RT [min]	W PeakID	W *m/z*	W RT [min]	W Deviation *m/z* [ppm]	W Deviation RT [min]
PE 34:1	C39H76NO8P	(-)	[M-H]-	716.52425	9.40	20287	716.52549	9.40	-1.74	0.00	7707	716.52523	9.40	-1.37	-0.01
PE 36:1	C41H80NO8P	(-)	[M-H]-	744.55543	10.46	21609	744.55703	10.45	-2.15	0.01	-	NA	NA	NA	NA
PE 36:2	C41H78NO8P	(-)	[M-H]-	742.53990	9.52	21510	742.54180	9.50	-2.57	0.02	16158	742.54101	9.65	-1.50	-0.13
PE 36:3	C41H76NO8P	(-)	[M-H]-	740.52425	8.77	21398	740.52546	8.80	-1.63	-0.03	16080	740.52551	8.80	-1.71	-0.03
PE 36:4	C41H74NO8P	(-)	[M-H]-	738.50860	8.02	21301	738.50907	8.02	-0.64	0.00	8124	738.50955	8.03	-1.29	-0.01
PE 36:5	C41H72NO8P	(-)	[M-H]-	736.49295	7.34	212061	736.49329	7.35	-0.47	-0.01	8090	736.49377	7.35	-1.12	-0.01
PE 36:6	C41H70NO8P	(-)	[M-H]-	734.47663	7.29	21095	734.47764	6.72	-1.38	0.57	-	NA	NA	NA	NA
PE 38:2	C43H82NO8P	(-)	[M-H]-	770.57120	10.72	22856	770.57276	10.72	-2.02	0.00	17110	770.57173	10.72	-0.69	0.00
PE 38:3	C43H80NO8P	(-)	[M-H]-	768.55555	9.77	22748	768.55559	9.76	-0.06	0.01	17020	768.56174	9.78	-8.06	-0.02
PE 38:4	C43H78NO8P	(-)	[M-H]-	766.53990	8.93	22641	766.54009	8.95	-0.26	-0.02	16921	766.54084	8.97	-1.23	-0.04
PE 40:2	C45H86NO8P	(-)	[M-H]-	798.60250	11.75	-	NA	NA	NA	NA	9803	798.60332	11.76	-1.03	-0.01
PE 40:3	C45H84NO8P	(-)	[M-H]-	796.58685	11.05	-	NA	NA	NA	NA	18108	796.58681	11.05	0.05	0.00
lysoPE 18:2a	C23H44NO7P	(-)	[M-H]-	476.27826	3.88	10864	476.27514	2.46	6.55	1.42	7920	476.27875	2.60	-1.04	1.28
lysoPE 18:2b	C23H44NO7P	(-)	[M-H]-	476.27826	4.04	10865	476.27844	2.75	-0.37	1.30	7921	476.27894	2.75	-1.42	1.29
lysoPE 18:3a	C23H42NO7P	(-)	[M-H]-	474.26261	3.41	10781	474.26314	2.09	-1.12	1.32	7855	474.26334	2.09	-1.54	1.32
lysoPE 18:3b	C23H42NO7P	(-)	[M-H]-	474.26261	3.58	10782	474.26309	2.23	-1.01	1.34	7856	474.26317	2.24	-1.17	1.34
PS 38:0	C44H86NO10P	(-)	[M-H]-	818.59166	9.22	-	NA	NA	NA	NA	19157	818.59242	9.24	-0.93	-0.02
PS 40:0	C46H90NO10P	(-)	[M-H]-	846.62296	10.28	-	NA	NA	NA	NA	20472	846.62360	10.33	-0.75	-0.05
SQDG 32:0	C41H78O12S	(-)	[M-H]-	793.51413	7.22	24091	793.51646	7.35	-2.94	-0.13	17969	793.51542	7.35	-1.63	-0.13
SQDG 32:2	C41H74O12S	(-)	[M-H]-	789.48283	-	23881	789.48524	6.28	-3.06	NA	-	NA	NA	NA	NA
SQDG 32:3	C41H72O12S	(-)	[M-H]-	787.46718	5.62	23760	787.48421	5.68	-21.63	-0.06	17748	787.46759	5.62	-0.53	0.00
SQDG 34:0	C43H82O12S	(-)	[M-H]-	821.54543	7.85	25740	821.54833	8.28	-3.53	-0.43	-	NA	NA	NA	NA
SQDG 34:1	C43H80O12S	(-)	[M-H]-	819.52978	7.40	25633	819.53168	7.57	-2.32	-0.17	-	NA	NA	NA	NA
SQDG 34:2	C43H78O12S	(-)	[M-H]-	817.51413	6.75	25494	817.51581	6.91	-2.06	-0.16	19085	817.51491	6.88	-0.96	-0.13
SQDG 34:3	C43H76O12S	(-)	[M-H]-	815.49848	6.23	25376	815.49937	6.35	-1.09	-0.12	18983	815.49932	6.33	-1.03	-0.10
SQDG 34:5	C43H72O12S	(-)	[M-H]-	811.46718	5.31	-	NA	NA	NA	NA	18767	811.46767	5.33	-0.60	-0.02
SQDG 36:1	C45H84O12S	(-)	[M-H]-	847.56108	8.34	27143	847.56287	8.51	-2.11	-0.17	-	NA	NA	NA	NA
SQDG 36:2	C45H82O12S	(-)	[M-H]-	845.54543	7.65	27031	845.54846	7.81	-3.59	-0.16	20433	845.55080	7.80	-6.36	-0.15
SQDG 36:3	C45H80O12S	(-)	[M-H]-	843.52978	6.98	26912	843.53009	7.18	-0.37	-0.20	20325	843.53077	7.18	-1.18	-0.20
SQDG 36:4 (1)	C45H78O12S	(-)	[M-H]-	841.51413	6.35	-	NA	NA	NA	NA	20202	841.51636	6.49	-2.65	-0.14
SQDG 36:4 (2)	C45H78O12S	(-)	[M-H]-	841.51413	6.41	26792	841.51634	6.50	-2.63	-0.09	20206	841.51237	6.57	2.09	-0.16
SQDG 36:5	C45H76O12S	(-)	[M-H]-	839.49848	5.87	26670	839.49953	5.97	-1.26	-0.10	-	NA	NA	NA	NA
SQDG 36:6	C45H74O12S	(-)	[M-H]-	837.48283	5.42	26563	837.48326	5.51	-0.51	-0.09	20002	837.48256	5.50	0.32	-0.08
TAG 48:0	C51H98O6	(+)	[M+NH4]+	824.76907	15.58	10166	824.76901	15.58	0.08	0.00	11515	824.76741	15.58	2.02	0.00
TAG 48:1	C51H96O6	(+)	[M+NH4]+	822.75358	15.34	10089	822.75329	15.34	0.35	0.00	11442	822.75281	15.34	0.93	0.00
TAG 48:2	C51H94O6	(+)	[M+NH4]+	820.73773	15.07	10023	820.73788	15.07	-0.19	0.00	11388	820.74041	15.07	-3.27	0.00
TAG 48:3	C51H92O6	(+)	[M+NH4]+	818.72252	14.74	9965	818.72219	14.74	0.40	0.00	11338	818.72000	14.73	3.08	0.01
TAG 48:4	C51H90O6	(+)	[M+NH4]+	816.70705	14.37	9904	816.70690	14.37	0.19	0.00	-	NA	NA	NA	NA
TAG 50:0	C53H102O6	(+)	[M+NH4]+	852.80124	15.84	11088	852.80129	15.84	-0.05	0.00	12360	852.79932	15.84	2.25	0.01
TAG 50:1	C53H100O6	(+)	[M+NH4]+	850.78356	15.60	11018	850.78435	15.60	-0.94	0.00	12287	850.78301	15.60	0.64	0.00
TAG 50:2	C53H98O6	(+)	[M+NH4]+	848.76826	15.35	10939	848.76846	15.35	-0.24	0.00	12197	848.76807	15.34	0.22	0.00
TAG 50:3	C53H96O6	(+)	[M+NH4]+	846.75297	15.08	10880	846.75280	15.08	0.20	0.00	12133	846.75143	15.07	1.82	0.00
TAG 50:4	C53H94O6	(+)	[M+NH4]+	844.73752	14.79	10836	844.73770	14.79	-0.22	0.00	12081	844.73622	14.78	1.53	0.01
TAG 50:5	C53H92O6	(+)	[M+NH4]+	842.72400	14.43	10780	842.72395	14.43	0.05	0.00	12040	842.72241	14.43	1.89	0.00
TAG 50:6	C53H90O6	(+)	[M+NH4]+	840.70664	14.07	10721	840.70725	14.07	-0.72	0.00	11980	840.70532	14.07	1.57	0.00
TAG 52:0	C55H106O6	(+)	[M+NH4]+	880.83325	16.08	12089	880.83306	16.08	0.21	0.00	13193	880.83108	16.08	2.46	0.00
TAG 52:1	C55H104O6	(+)	[M+NH4]+	878.81556	15.85	12009	878.81630	15.85	-0.84	0.00	13121	878.81455	15.85	1.15	0.01
TAG 52:2	C55H102O6	(+)	[M+NH4]+	876.79945	15.62	11921	876.79960	15.62	-0.17	0.00	13055	876.79844	15.63	1.16	0.00
TAG 52:3	C55H100O6	(+)	[M+NH4]+	874.78355	15.38	11830	874.78387	15.38	-0.37	0.00	12985	874.78258	15.38	1.10	0.00
TAG 52:4	C55H98O6	(+)	[M+NH4]+	872.76789	15.10	11752	872.76775	15.10	0.16	0.00	12917	872.76614	15.11	2.01	-0.01
TAG 52:5	C55H96O6	(+)	[M+NH4]+	870.75137	14.81	11673	870.75192	14.81	-0.64	0.00	12858	870.75064	14.81	0.84	0.00
TAG 52:6	C55H94O6	(+)	[M+NH4]+	868.73553	14.50	11616	868.73590	14.50	-0.43	0.00	12806	868.73469	14.50	0.97	0.00
TAG 52:7	C55H92O6	(+)	[M+NH4]+	866.72230	14.16	11557	866.72188	14.16	0.48	0.00	12751	866.72067	14.15	1.88	0.01
TAG 52:8	C55H90O6	(+)	[M+NH4]+	864.70734	13.75	11487	864.70710	13.76	0.28	0.00	12682	864.70414	13.75	3.71	0.00
TAG 52:9	C55H88O6	(+)	[M+NH4]+	862.68979	13.28	11403	862.68984	13.28	-0.06	0.00	-	NA	NA	NA	NA
TAG 54:0	C57H110O6	(+)	[M+NH4]+	908.86370	16.31	13083	908.86400	16.31	-0.33	0.00	14086	908.86222	16.31	1.63	0.00
TAG 54:1	C57H108O6	(+)	[M+NH4]+	906.84757	16.09	12998	906.84844	16.09	-0.95	0.00	13998	906.84631	16.08	1.40	0.01
TAG 54:2	C57H106O6	(+)	[M+NH4]+	904.83186	15.87	12920	904.83198	15.87	-0.12	0.00	13926	904.83033	15.87	1.70	0.00
TAG 54:3	C57H104O6	(+)	[M+NH4]+	902.81503	15.65	12840	902.81497	15.65	0.07	0.00	13859	902.81295	15.65	2.31	0.00
TAG 54:4	C57H102O6	(+)	[M+NH4]+	900.79863	15.40	12767	900.79935	15.40	-0.80	0.00	13798	900.79766	15.41	1.08	-0.01
TAG 54:5	C57H100O6	(+)	[M+NH4]+	898.78281	15.14	12708	898.78399	15.14	-1.32	0.00	13752	898.78265	15.15	0.17	-0.01
TAG 54:6	C57H98O6	(+)	[M+NH4]+	896.76724	14.85	12645	896.76755	14.85	-0.35	0.00	13695	896.76636	14.86	0.98	-0.01
TAG 54:7	C57H96O6	(+)	[M+NH4]+	894.75216	14.54	12570	894.75095	14.54	1.36	0.00	13627	894.75195	14.55	0.24	-0.01
TAG 54:8	C57H94O6	(+)	[M+NH4]+	892.73617	14.19	12483	892.73576	14.20	0.46	0.00	13547	892.73549	14.20	0.77	0.00
TAG 56:1	C59H112O6	(+)	[M+NH4]+	934.87898	16.33	14007	934.87929	16.34	-0.33	0.00	14908	934.87881	16.32	0.18	0.01
TAG 56:2	C59H110O6	(+)	[M+NH4]+	932.86267	16.11	13935	932.86333	16.11	-0.70	0.00	148371	932.86227	16.10	0.44	0.01
TAG 56:3	C59H108O6	(+)	[M+NH4]+	930.84753	15.89	13872	930.84745	15.89	0.08	0.00	14780	930.84566	15.88	2.00	0.00
TAG 56:4	C59H106O6	(+)	[M+NH4]+	928.83116	15.65	13808	928.83094	15.66	0.23	0.00	14723	928.82902	15.66	2.30	0.00
TAG 56:5	C59H104O6	(+)	[M+NH4]+	926.81543	15.41	13743	926.81567	15.41	-0.26	0.00	14671	926.81316	15.41	2.45	0.00
TAG 56:6	C59H102O6	(+)	[M+NH4]+	924.79787	15.15	13666	924.79914	15.15	-1.37	0.00	-	NA	NA	NA	NA
TAG 56:8	C59H98O6	(+)	[M+NH4]+	920.76529	14.20	13479	920.76573	14.20	-0.48	0.00	14443	920.76285	14.21	2.64	-0.01
TAG 56:9	C59H96O6	(+)	[M+NH4]+	918.75831	14.25	13407	918.75852	14.25	-0.23	0.00	14381	918.75117	14.25	7.77	-0.01
TAG 58:1	C61H116O6	(+)	[M+NH4]+	962.91057	16.57	15031	962.91048	16.57	0.09	0.00	-	NA	NA	NA	NA
TAG 58:2	C61H114O6	(+)	[M+NH4]+	960.89493	16.34	14957	960.89494	16.34	-0.01	0.00	15796	960.89345	16.33	1.55	0.01
TAG 58:3	C61H112O6	(+)	[M+NH4]+	958.87927	16.12	14895	958.87901	16.12	0.27	0.00	15738	958.87772	16.11	1.61	0.01
TAG 58:4	C61H110O6	(+)	[M+NH4]+	956.86317	15.91	14831	956.86319	15.91	-0.02	0.00	15673	956.86089	15.90	2.38	0.01
TAG 58:5	C61H108O6	(+)	[M+NH4]+	954.84792	15.69	14759	954.84761	15.69	0.33	0.00	15603	954.84491	15.67	3.15	0.02
TAG 60:1	C63H120O6	(+)	[M+NH4]+	990.94176	16.80	15983	990.94220	16.80	-0.44	0.00	-	NA	NA	NA	NA
TAG 60:3	C63H116O6	(+)	[M+NH4]+	986.91055	16.34	15866	986.91111	16.34	-0.57	0.00	16592	986.90923	16.33	1.34	0.01
TAG 60:4	C63H114O6	(+)	[M+NH4]+	984.89396	16.13	15805	984.89375	16.13	0.22	0.00	-	NA	NA	NA	NA
TAG 60:5	C63H112O6	(+)	[M+NH4]+	982.87767	15.93	15729	982.87786	15.93	-0.19	0.00	-	NA	NA	NA	NA



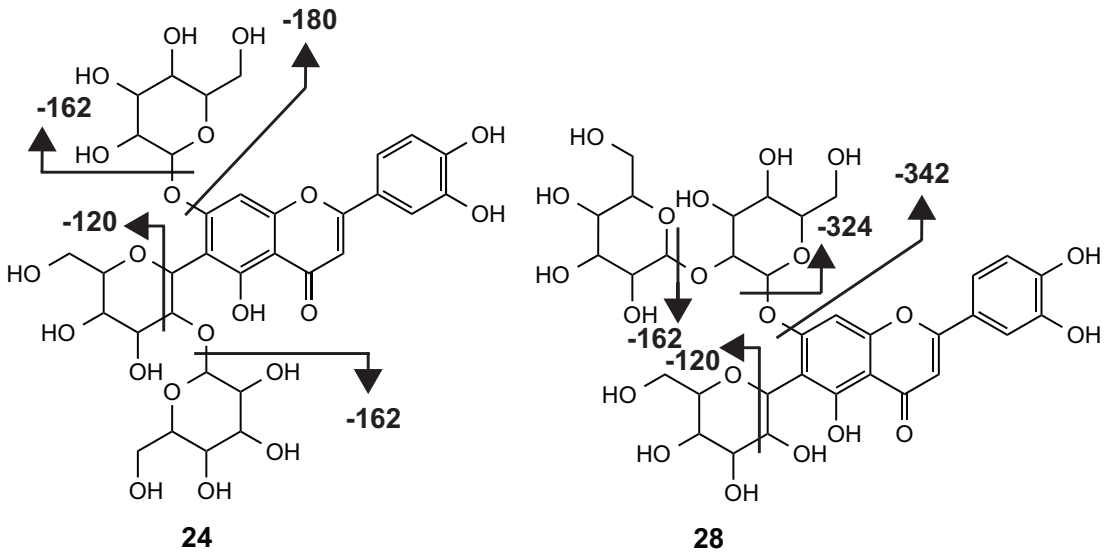
Supplementary figure 16: Example of total ion chromatogram (*m/z* 150–1500) of lipid analysis of one replicate of BH2 Ctrl low with polarity switch after 13 min.

Supplementary table 6: Component list of putatively identified secondary metabolites in barely and rice based on intersection of high-expressing genotypes. Reference: Coutinho [1]

Component	Formula	Ion	Peak *m/z*	Deviation *m/z* [ppm]	Peak RT [min]	Intensity (BH low) Ctrl	p-value	FC	B	R	Putative annotation	Annotation level	Reference
o	$C_{13}H_{16}O_9$	$[C_6H_4O_2]^-$	108.02025	-13.2	3.65	831	2.0E-14	26.72	X		gentisic acid fragment	2	Coutinho
		$[C_7H_4O_4]^-$	152.01021	-8.6	3.65	1821	4.0E-14	26.60			gentisic acid fragment		
		$[C_6H_{11}O_6]^-$	179.05509	-5.7	3.66	3663	1.9E-09	231.53			glucose fragment		
		$[M-H]^-$	315.07214	-0.1	3.65	133860	8.3E-14	24.88			gentisic acid *O*-glucoside		
p	$C_{23}H_{38}O_{20}$	$[M-H]^-$	633.18900	1.0	2.38	230450	5.3E-01	0.92		X	Tetrasaccharide	–	–
		$[M+NH_4]^+$	652.22982	0.5	2.39	6718	3.7E-03	0.61					
		$[M+Na]^+$	657.18580	1.4	2.39	44593	1.6E-05	0.60					
		$[M+K]^+$	673.15900	0.3	2.39	6912	8.4E-01	1.07					
q	$C_{24}H_{40}O_{20}$	$[M-H]^-$	647.20427	0.4	2.87	338429	8.6E-01	1.06		X	*O*-methoxylated tetrasaccharide	–	
		$[M+H]^+$	649.21926	1.1	2.87	67364	2.5E-04	0.72					
		$[M+NH_4]^+$	666.24335	-2.7	2.87	9596	1.2E-03	0.58					
		$[M+Na]^+$	671.20120	1.0	2.87	64482	8.7E-06	0.62					
		$[M+K]^+$	687.17465	0.3	2.87	9104	5.5E-01	0.88					

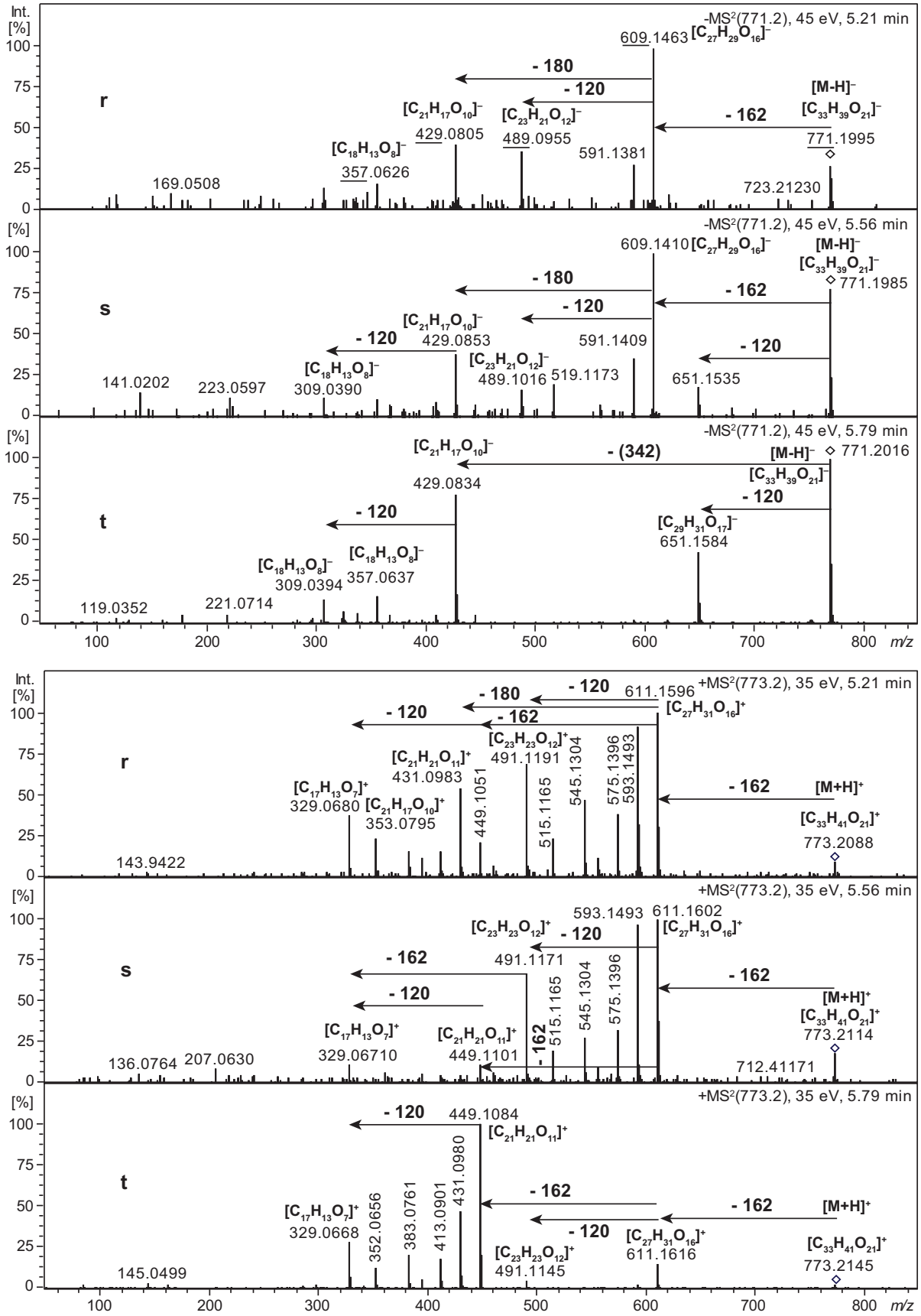


Supplementary figure 17: MS/MS spectra of component **p** and **q** in (–)-ESI on top and (+)-ESI on the bottom.

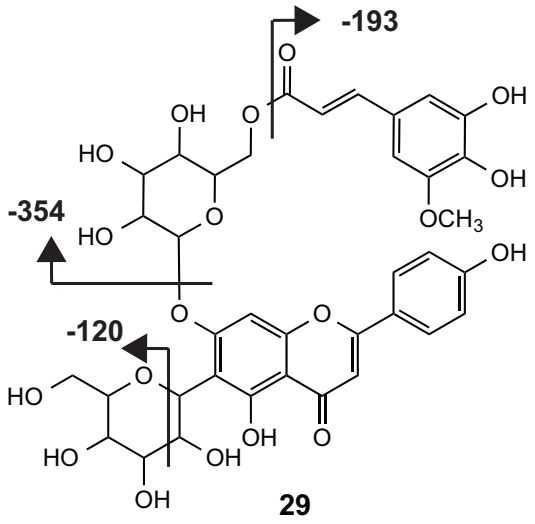


24 **28**

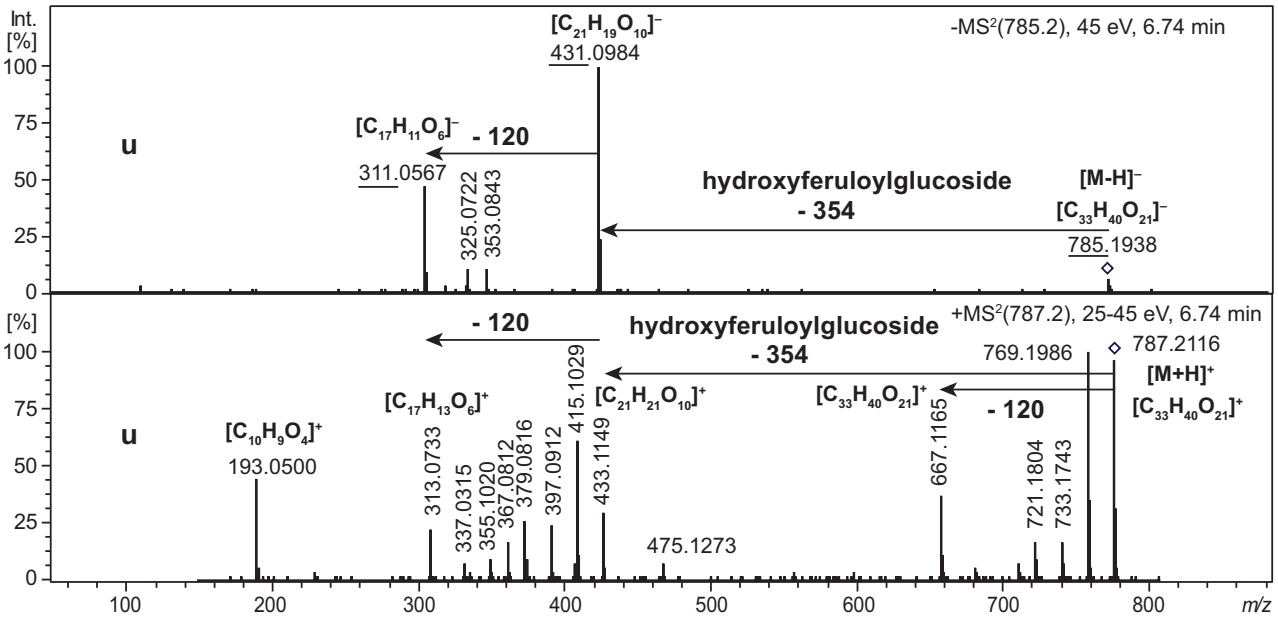
Supplementary scheme 2: Putative annotation of component **r, s** and **t**, $C_{27}H_{30}O_{15}$: isoorientin-7, 2”-di-*O*-glucoside (**24**) reported in barley [2], which was putatively assigned to component **r**, and isoorientin-7-*O*-diglucoside (**28**) that was putatively assigned to component **t**. Component **s** was putatively assigned as isomer of (**24**), isoorientin-di-*O*-hexoside, but the exact position of the *O*-glycosylations could not be determined.



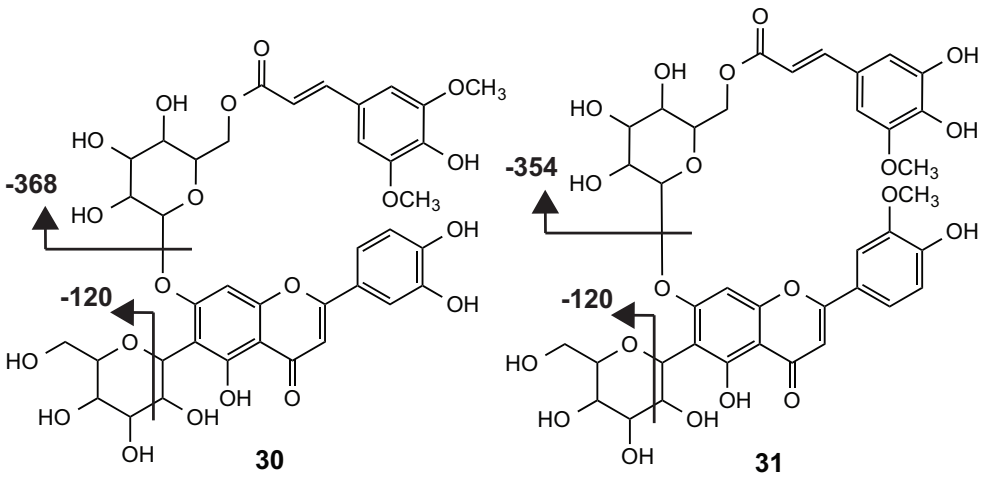
Supplementary figure 18: MS/MS spectra of components **r**, **s** and **t** ($C_{33}H_{40}O_{21}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments reported in [2] (–).



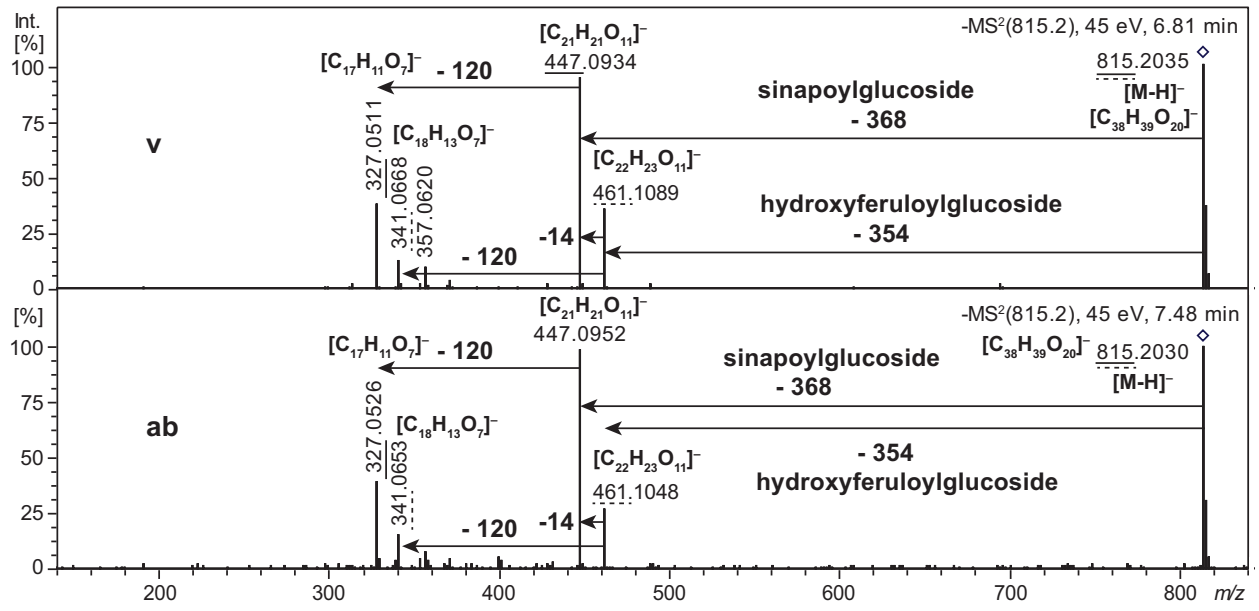
Supplementary scheme 3: Putative annotation of component **u**, $C_{37}H_{38}O_{19}$: isovitexin-7-*O*-[6”-hydroxyferuloyl]-glucoside (**29**) reported in barley [2].



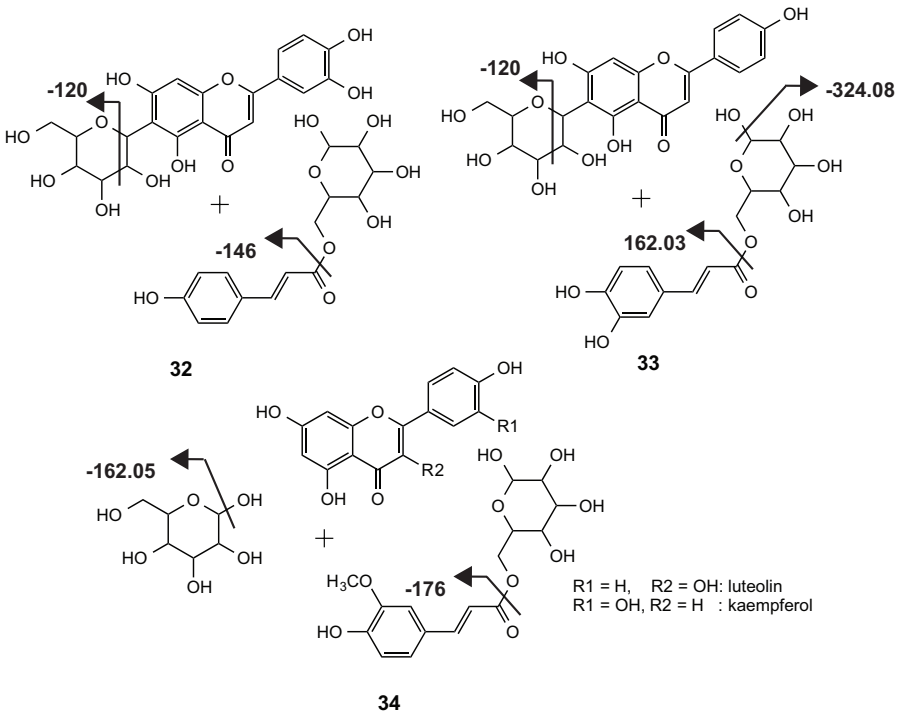
Supplementary figure 19: MS/MS spectra of component **u** ($C_{37}H_{38}O_{19}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments reported in [2] (–).



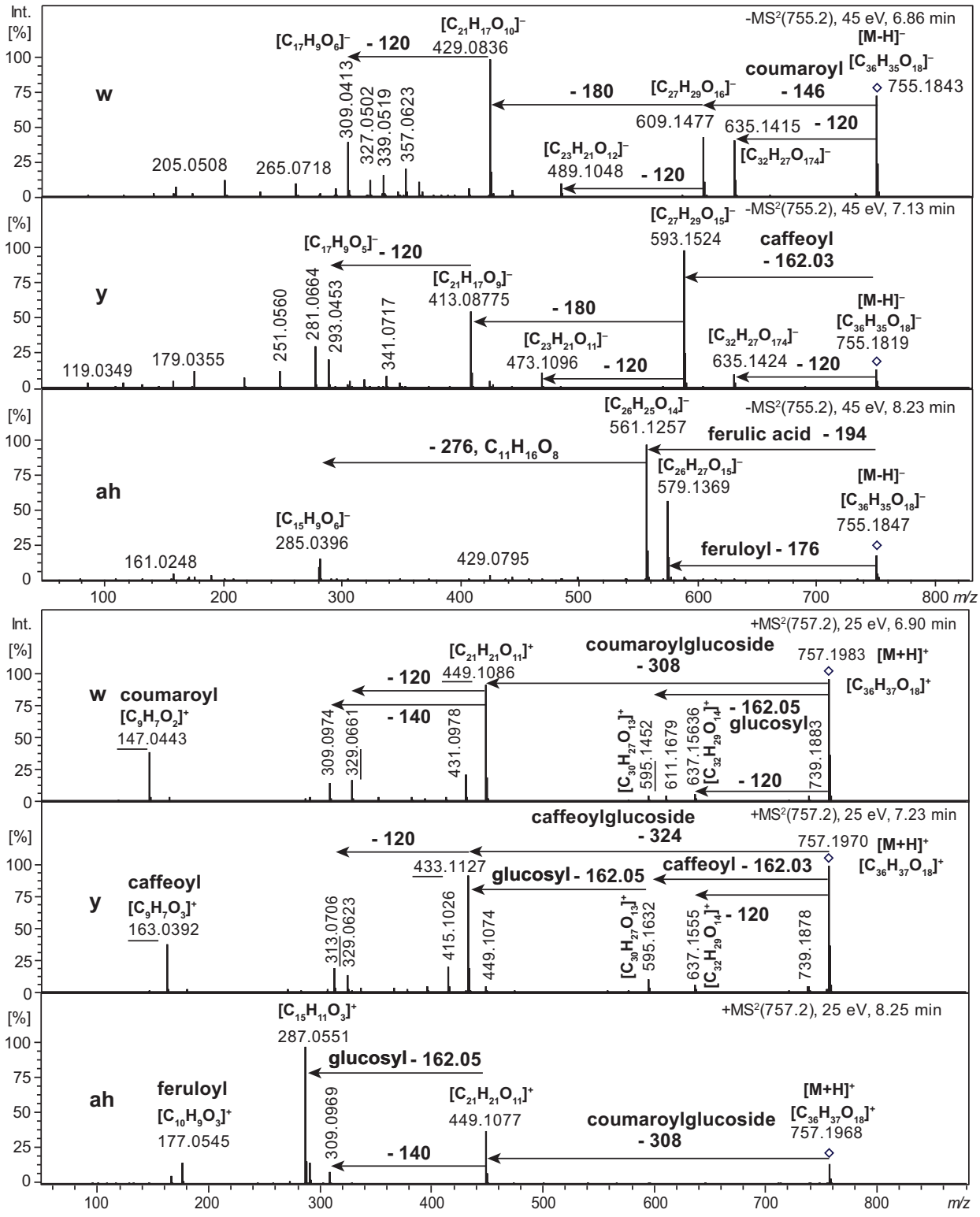
Supplementary scheme 4: Putative annotation of component **v** and **ab**, $C_{38}H_{40}O_{20}$: mixture between isoorientin-7-*O*-[6"-hydroxyferuloyl]glucoside (**30**) and isoscoparin-7-*O*-[6"-sinypoyl]-glucoside (**31**) reported in barley [2].



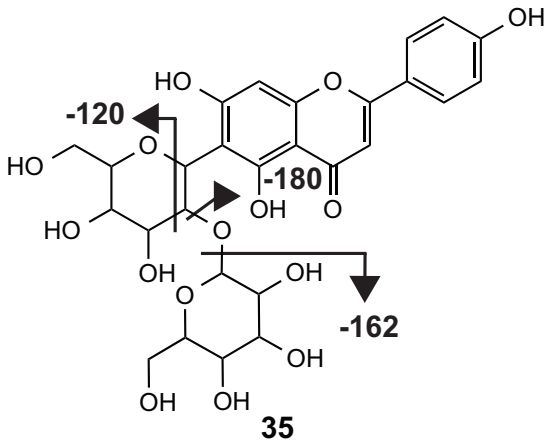
Supplementary figure 20: MS/MS spectra of components **v** and **ab** ($C_{38}H_{40}O_{20}$) in (–)-ESI, underlined fragments reported for **30** (dashed line) and for **31** (normal line) [2].



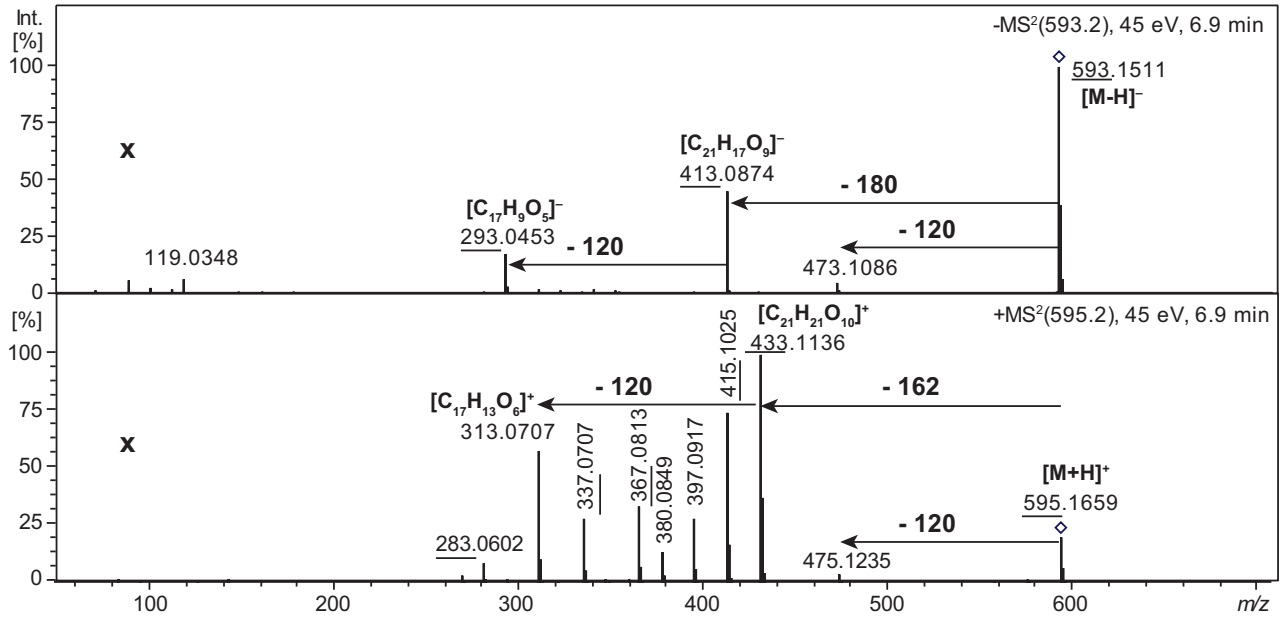
Supplementary scheme 5: Putative annotation of components **w**, **y**, and **ah**, $C_{36}H_{36}O_{18}$: component **w** assigned as isomer of *C*-hexosyl-luteolin-*p*-coumaroylhexoside, illustrated as isoorientin-*p*-coumaroyl-glucoside (**32**), component **y** as *C*-hexosyl-apigenin-caffeoylhexoside, illustrated as isovitexin-*p*-coumaroylglucoside (**33**), reported in rice in [3, 4] and component **ah** as luteolin- or kaempferol-feruloylhexoside (**34**). The exact position of the *C*- and *O*-glycosylations as well as for the acetylations could not be determined.



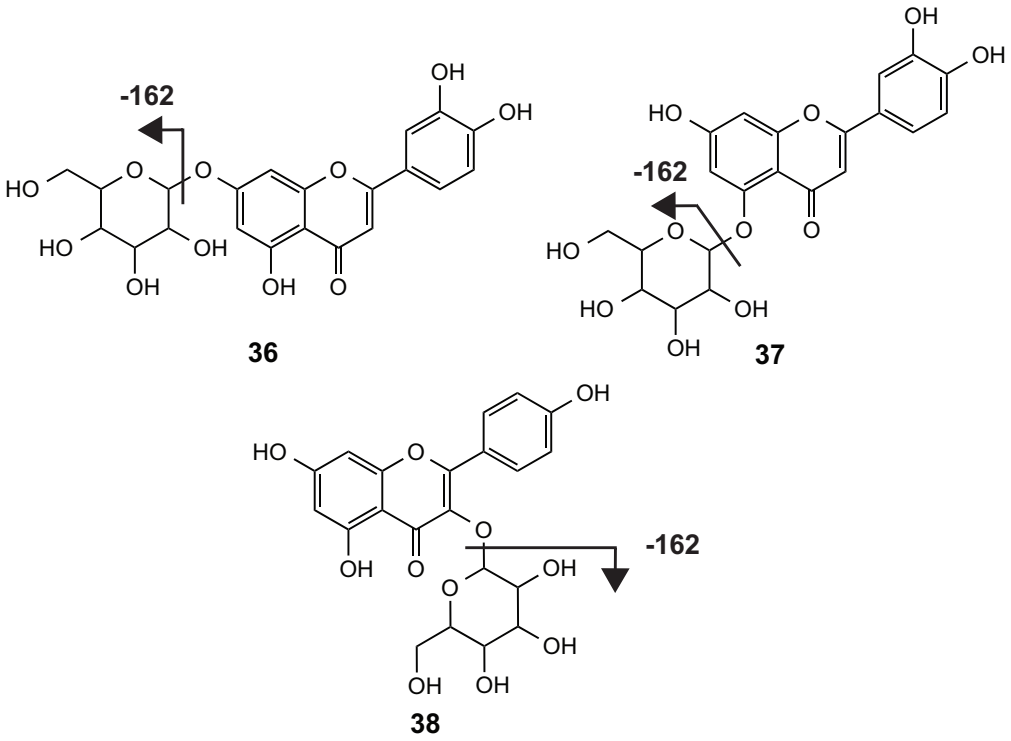
Supplementary figure 21: MS/MS spectra of components **w**, **y** and **ah** ($C_{36}H_{36}O_{18}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments reported in [3, 4] (+), additional fragments were reported for **32** (*m/z* 395), and for **33** (*m/z* 366, 337, 283).



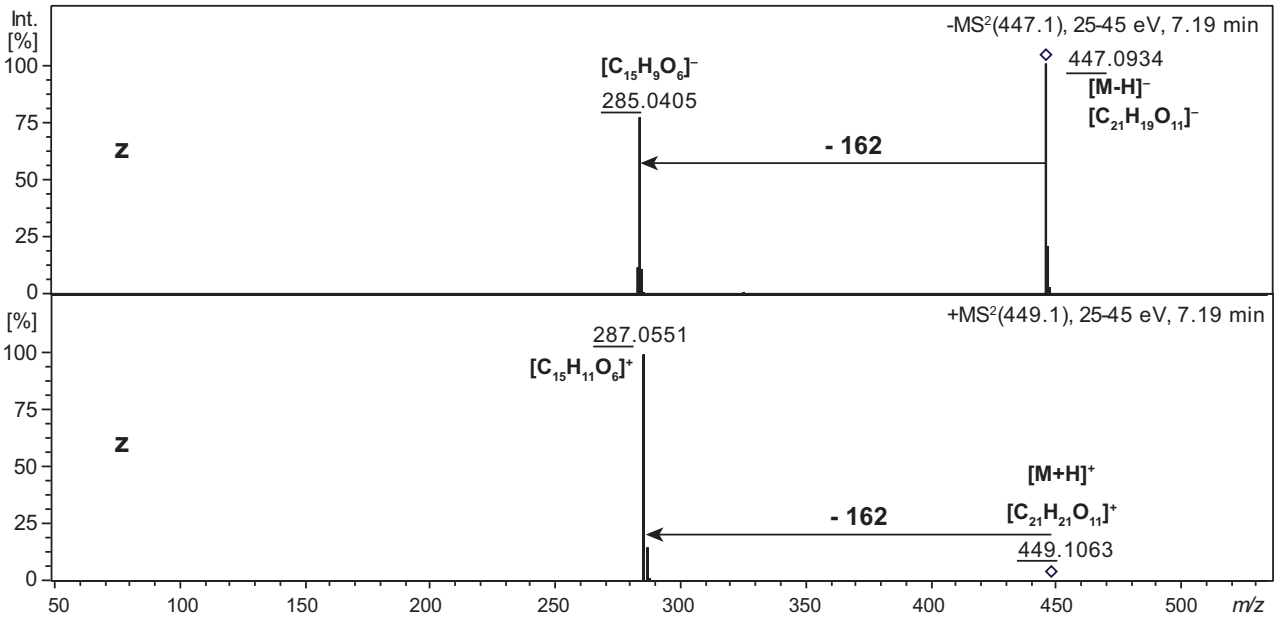
Supplementary scheme 6: Putative annotation of component **x**, $C_{27}H_{30}O_{15}$: isovitexin-2”-*O*-glucoside **35** reported in barley [2].



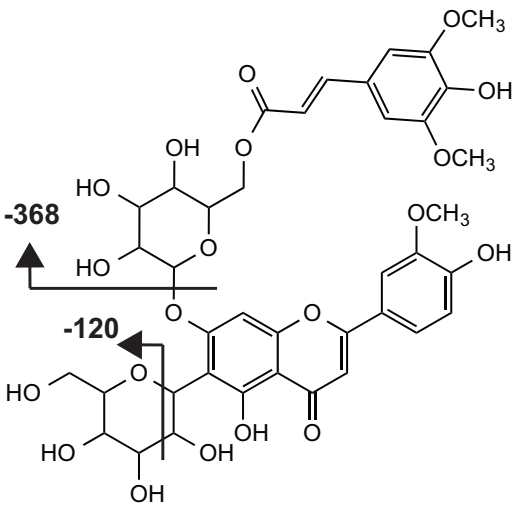
Supplementary figure 22: MS/MS spectra of component **x** ($C_{27}H_{30}O_{15}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments reported in [2] (+), (–).



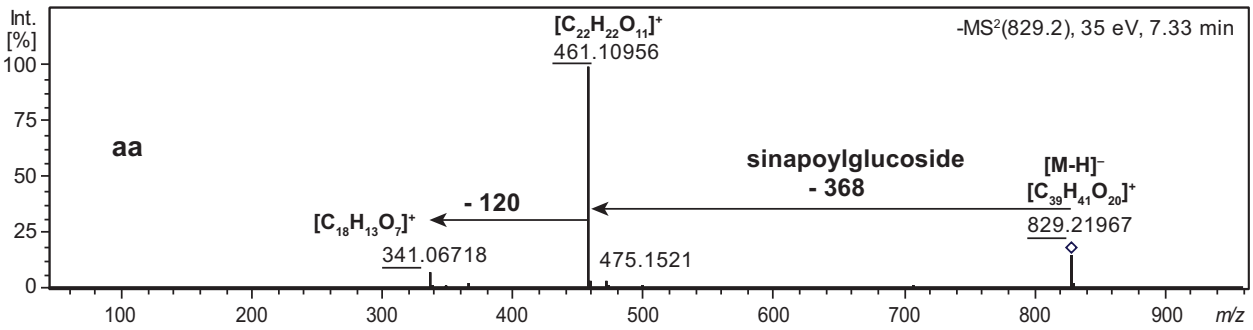
Supplementary scheme 7: Putative annotation of component **z**, $C_{21}H_{20}O_{11}$: luteolin-7-*O*-glucoside (**36)**, luteolin-5-*O*-glucoside (**37**) or kaempferol 3-*O*-glucoside (**38**) reported in rice [3, 4]



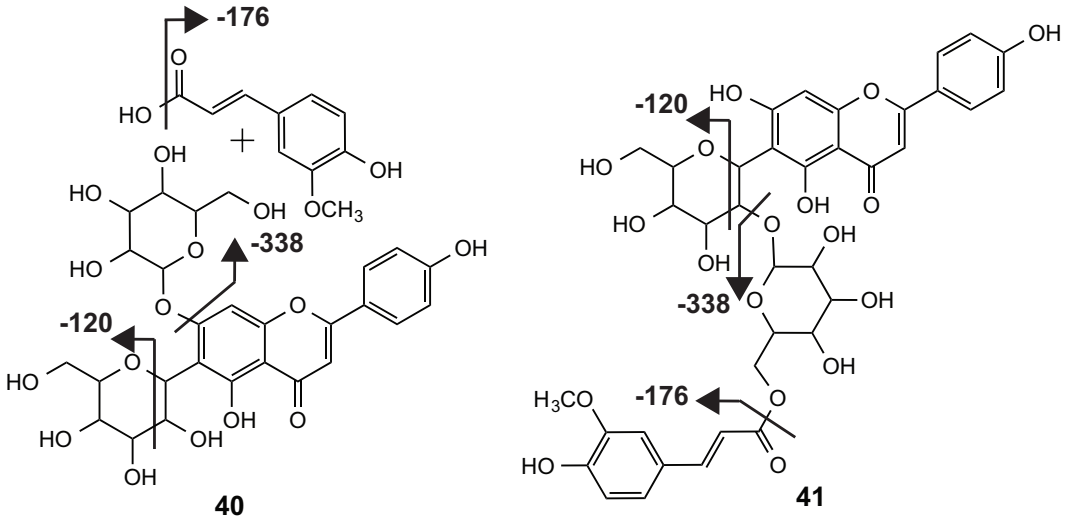
Supplementary figure 23: MS/MS spectra of component **z** ($C_{21}H_{20}O_{11}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments reported in [2] (+), (–). In [4] and [3] (+), additional fragments were reported for **36** (*m/z* 413, 353, 329, 299), for **37** (*m/z* 413, 329) and for **38** (*m/z* 413, 383, 329, 299).



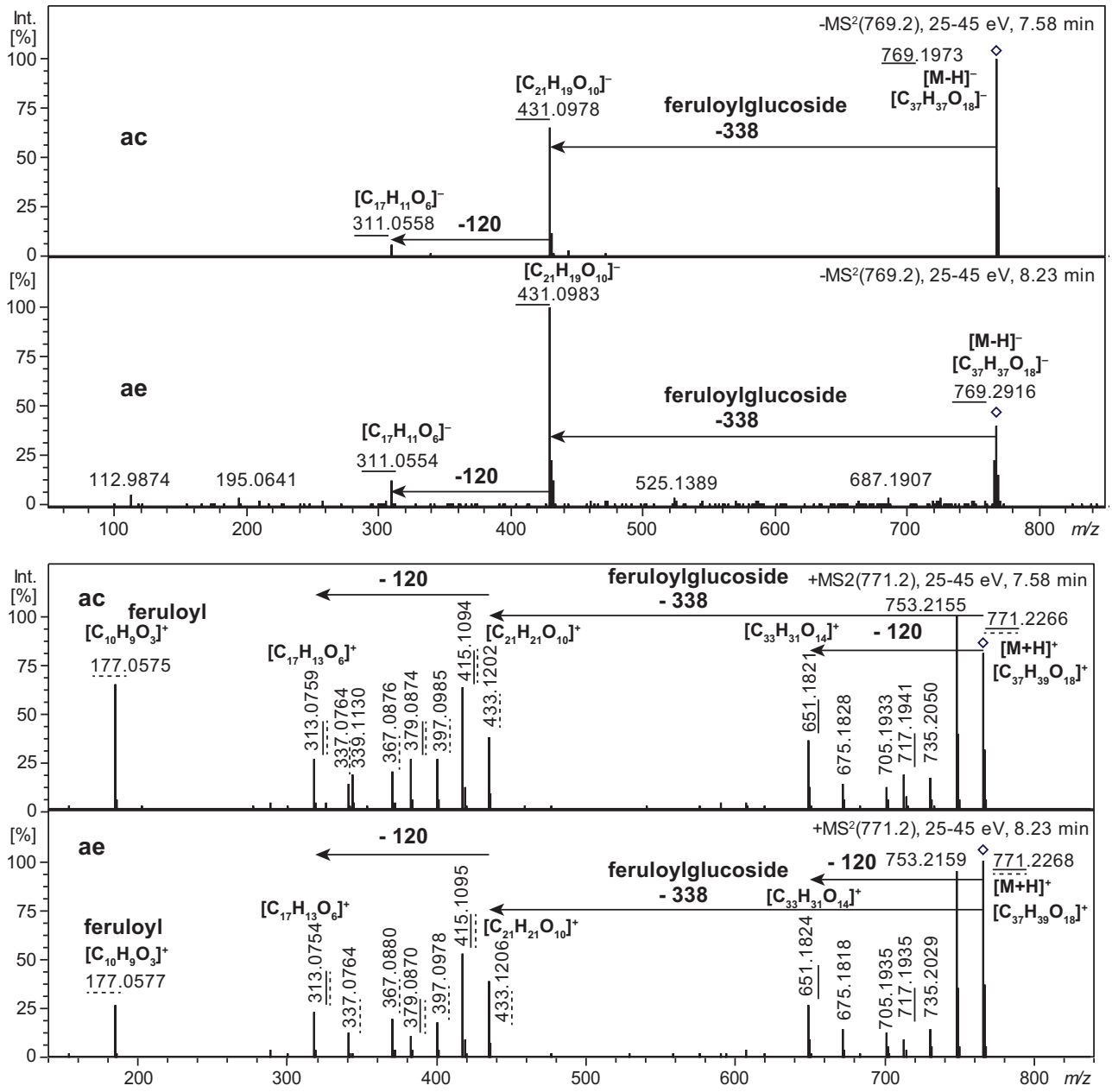
Supplementary scheme 8: Putative annotation of component **aa**, $C_{38}H_{42}O_{20}$: Isoscoparin-7-*O*-[6”-sinapoyl]glucoside (**39**), reported in barley [2].



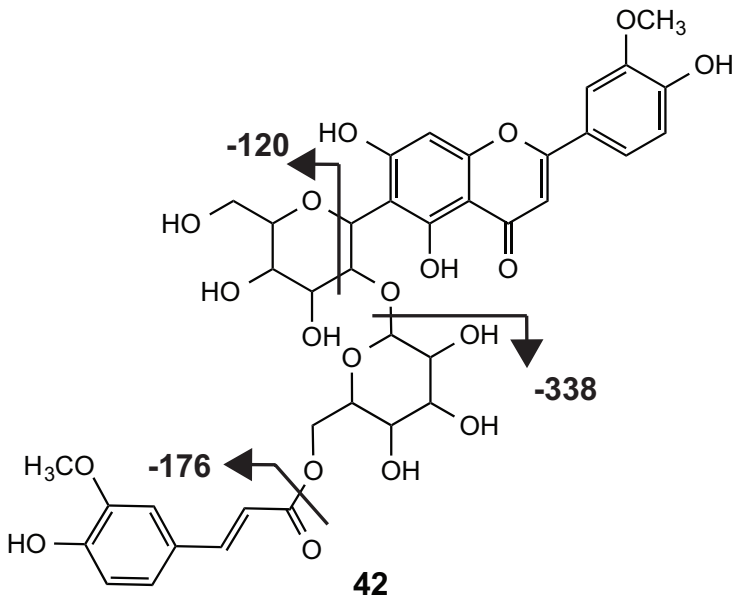
Supplementary figure 24: MS/MS spectra of component **aa** ($C_{39}H_{42}O_{20}$) in (–)-ESI, underlined fragments were reported in [2] (–).



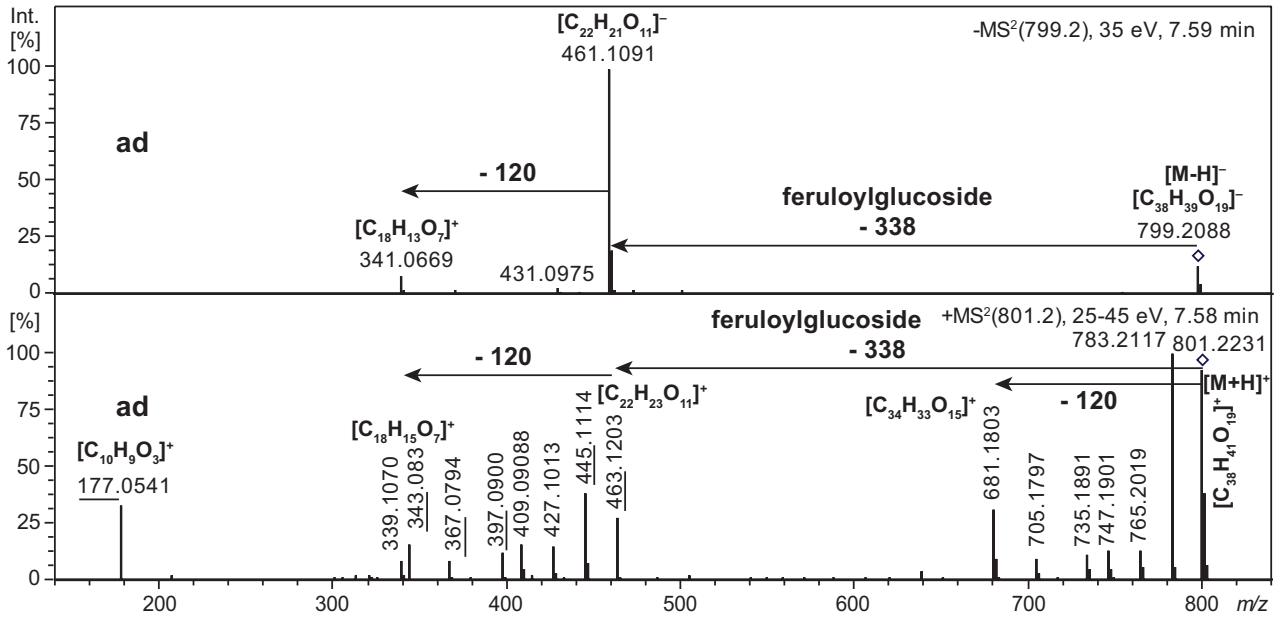
Supplementary scheme 9: Putative annotation of components **ac** and **ae**, $C_{37}H_{38}O_{18}$: isomers of isovitexin-7-*O*-[X”-feruloyl]-glucoside (**40**) reported in barley [2] or isovitexin-2”-*O*-[6”-(*E*)-feruloyl]glucoside (**41**) reported in rice [5]. The exact position of the *O*-glycosylations and acetylations could not be determined.



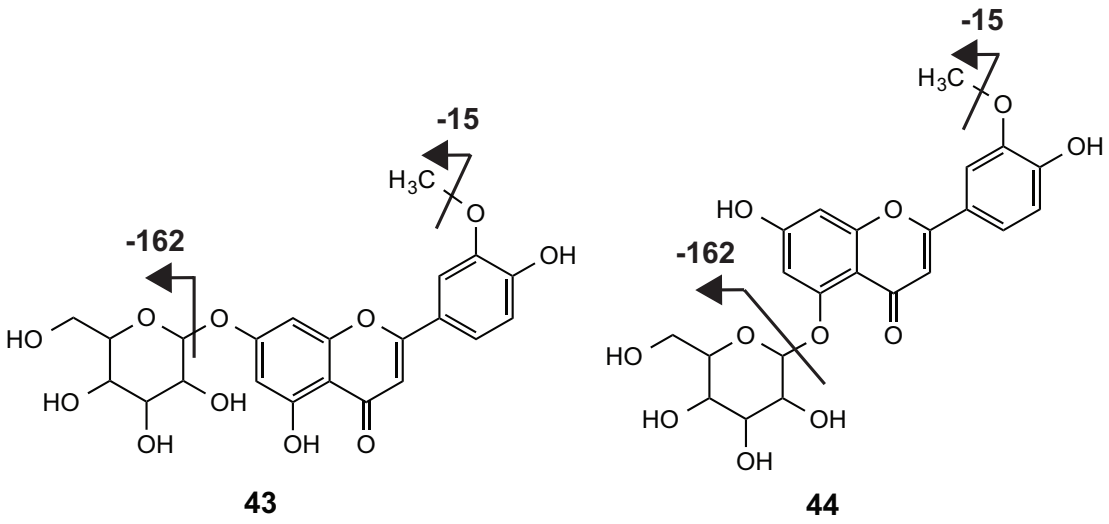
Supplementary figure 25: MS/MS spectra of components **ac** and **ae** ($C_{37}H_{38}O_{18}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments were reported in [2](+), (–) (normal line), and in [5] (+)(dashed line). Additional fragments were reported for **40** (*m/z* 609, 523, 361, 283).



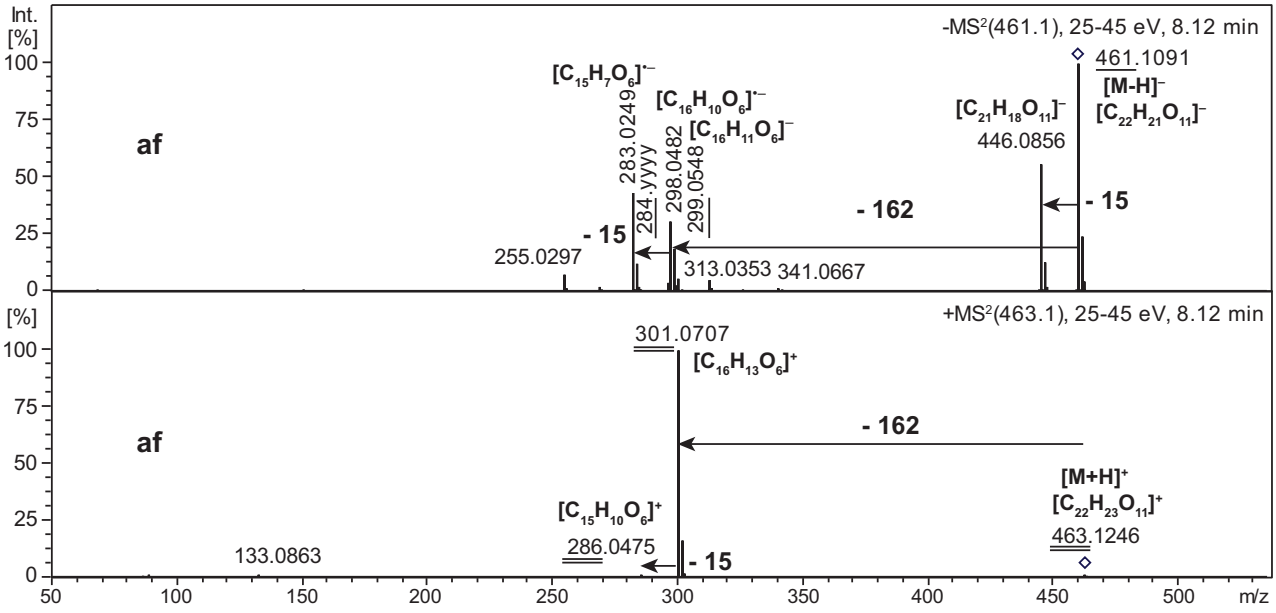
Supplementary scheme 10: Putative annotation of components **ad**, $C_{38}H_{40}O_{19}$: isoscoparin 2”-O-(6”-(E)-feruloyl)-glucopyranoside (**42**) was reported in rice [5].



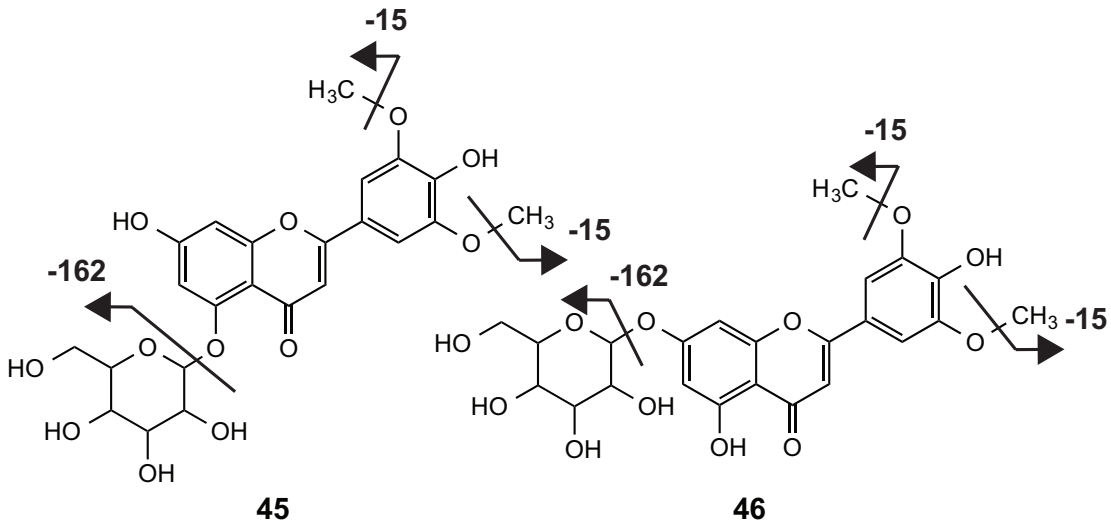
Supplementary figure 26: MS/MS spectra of component **ad** ($C_{38}H_{40}O_{19}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments were reported in references [5](+).



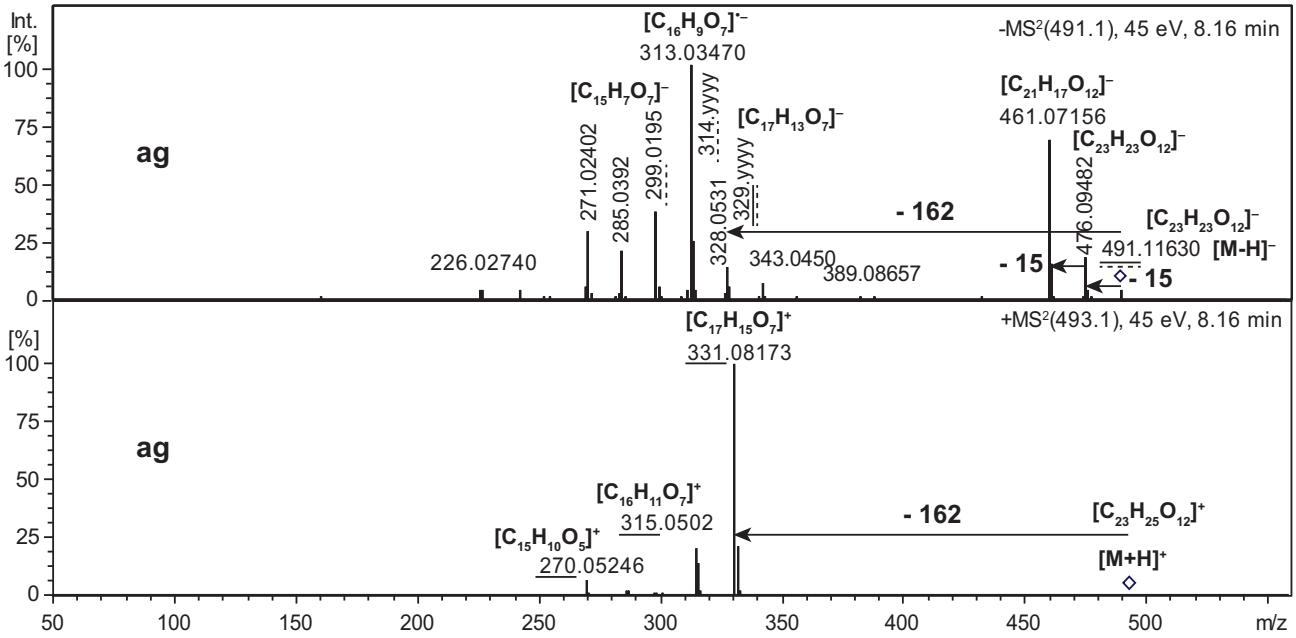
Supplementary scheme 11: Putative annotation of component **af**, $C_{22}H_{22}O_{11}$: chrysoeriol-7-*O*-glucoside (**43**) reported in barley [2] and rice [3] and chrysoeriol-5-*O*-glucoside (**44**) reported in rice [3].



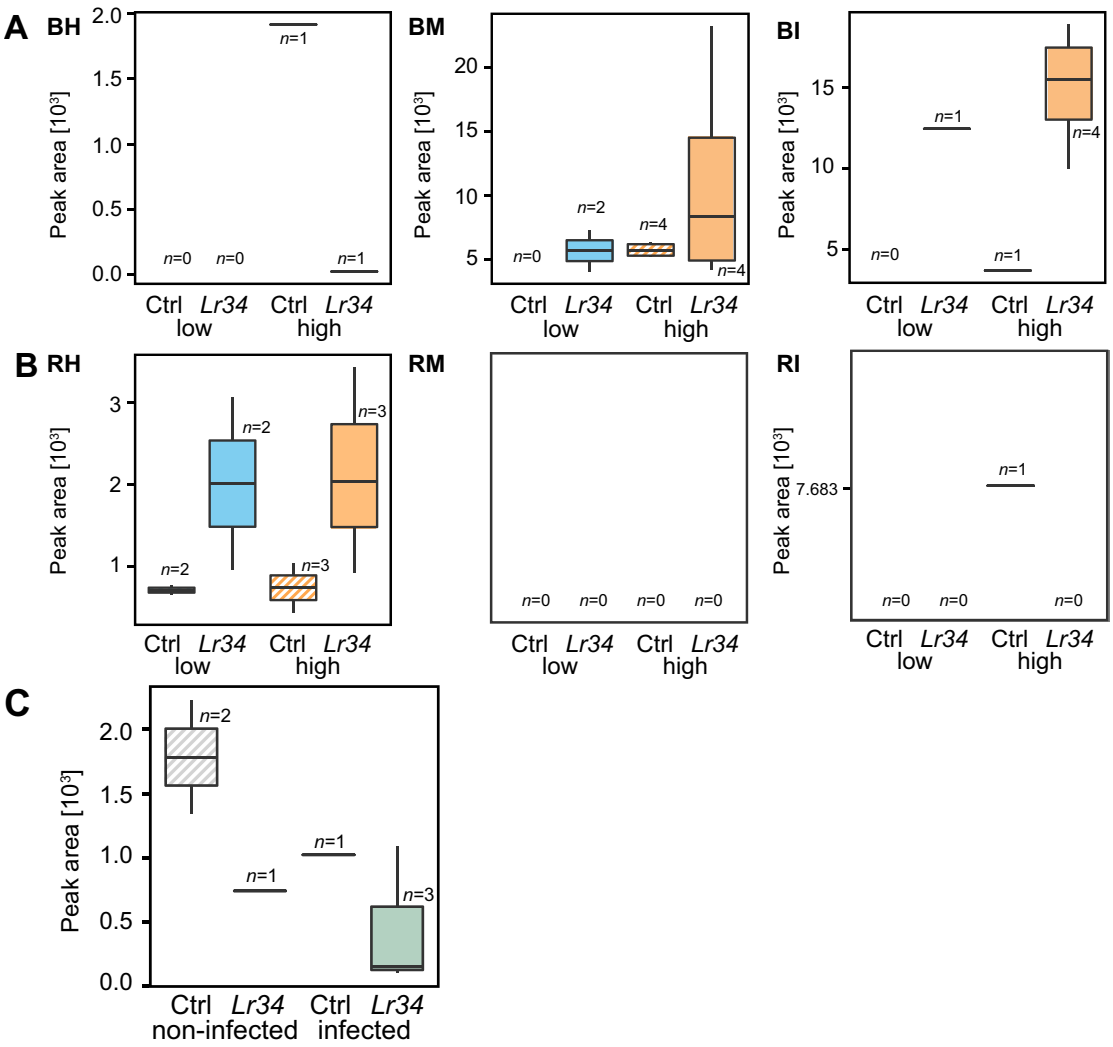
Supplementary figure 27: MS/MS spectra of component **af** ($C_{22}H_{22}O_{11}$) in (–)-ESI on top and (+)-ESI on the bottom, underlined fragments were reported in references [2](+) and (–), [4] (+); no difference between the spectra of **43** and **44** was reported.



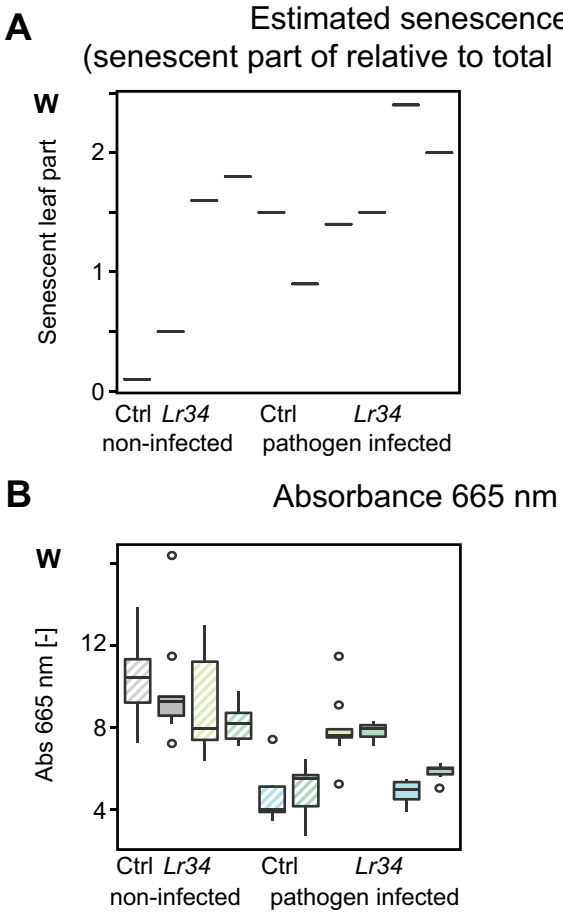
Supplementary scheme 12: Putative annotation of component **ag**, $C_{23}H_{24}O_{12}$: tricin-5-*O*-glucoside (**45**) reported in rice [3-6] or tricin-7-*O*-glucoside (**46**) reported in barley [2] and in rice [4-6].



Supplementary figure 28: MS/MS spectra of component **ag** ($C_{23}H_{24}O_{12}$) in (–)-ESI on top and (+)-ESI on the bottom. Underlined fragments were reported in (+)-ESI in [2-6] for both **45** and **46**, in (–)-ESI in [5] (dashed line) and [2] (normal line).



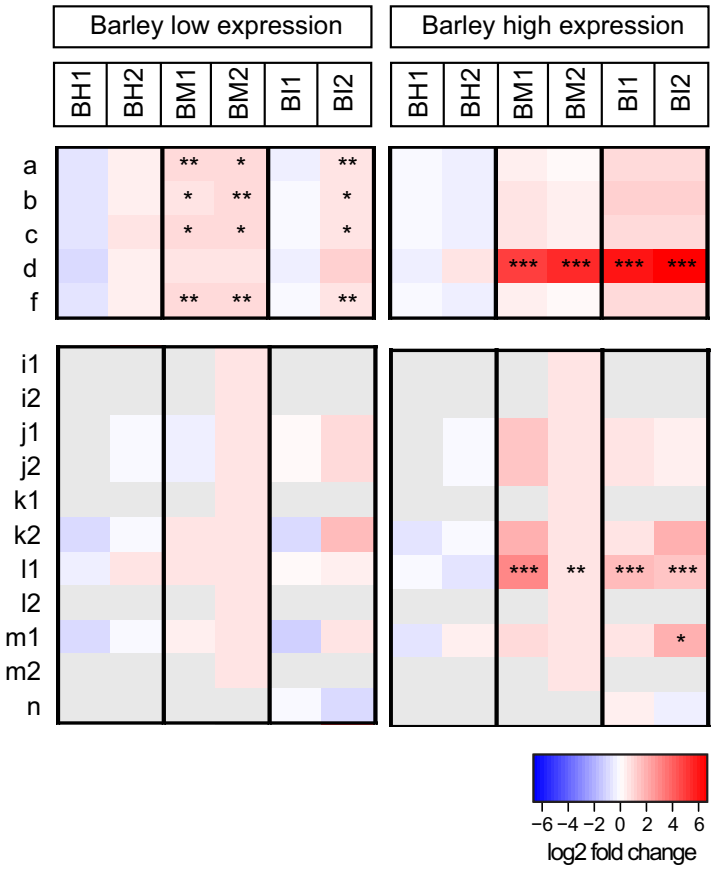
Supplementary figure 29: Signals of JA (**26**) in different experiments of barley (A), rice (B) and wheat (C).



Supplementary figure 30: Evaluation of senescence in wheat: estimated senescence based on the senescence part of the harvested leaf relative to the leaf length scaled to 10 (A), and measurement of chlorophyll A based on absorbance at 665 nm (B).

Supplementary table 7: Components of previously annotated secondary metabolites including cyanoglycoside derivatives (**a**-**d** and **f**) and hordatine derivatives (**i**–**n**) in barley detected in metabolic fingerprinting (0.4)

Com-ponent	Putative annotation	Formula	Ion	PeakID	Peak *m/z*	Deviation *m/z* [ppm]	Peak RT [min]	Intensity BH Ctrl low	*p*-value BM2 low	FC BM2 low
a	Sutherlandin	C11H17NO6	$[M+HCOO]^-$	49550	320.09858	-0.4	2.22	369961	7.3E-03	1.79
			$[M+H]^+$	51101	276.10780	0.1	2.23	1441717	2.6E-02	1.16
			$[M+NH_4]^+$	56027	293.13416	-0.6	2.22	105328	4.7E-01	0.85
			$[M+Na]^+$	57311	298.08961	-0.4	2.23	1148773	5.0E-01	0.63
b	Epidermin	C11H19NO6	$[M+HCOO]^-$	46421	306.11924	-0.5	3.21	1608882	1.7E-02	2.15
			$[M+NH_4]^+$	51997	279.15526	0.8	3.22	545716	3.3E-02	1.05
			$[M+Na]^+$	53271	284.11029	-0.5	3.22	2938827	1.0E-02	0.91
c	Osmaronin	C11H17NO6	[M-H]-	36090	258.09801	-1.2	3.49	20854	6.0E-03	2.13
			$[M+HCOO]^-$	45892	304.10341	-1.3	3.49	393512	1.4E-02	1.98
			$[M+H]^+$	46608	260.11294	0.2	3.50	1120529	4.9E-03	1.65
			$[M+NH_4]^+$	51440	277.13973	1.1	3.50	95018	1.9E-02	1.15
			$[M+Na]^+$	52664	282.09475	-0.2	3.50	1956284	1.9E-03	1.07
d	Dihydroosmaronin	C11H19NO6	[M-H]-	36459	260.11358	-1.3	3.61	12370	8.9E-03	2.03
			$[M+HCOO]^-$	46422	306.11933	-0.2	3.65	44897	9.6E-01	101.45
			$[M+H]^+$	47240	262.12860	0.4	3.61	66683	5.0E-03	1.20
			$[M+NH_4]^+$	51998	279.15457	-1.6	3.61	11956	5.6E-02	1.39
			$[M+Na]^+$	53274	284.11049	0.2	3.61	219212	1.5E-02	0.73
f	Epiheterodendrin	C11H19NO6	$[M+HCOO]^-$	46423	306.11918	-0.7	4.66	1040116	7.1E-03	1.74
			$[M+H]^+$	47242	262.12868	0.8	4.68	13976	5.4E-03	1.52
			$[M+NH_4]^+$	51996	279.15489	-0.5	4.68	201667	3.9E-02	0.95
			$[M+Na]^+$	53275	284.11024	-0.6	4.68	1169414	1.8E-02	0.59
i1	Hordatine A glucoside	C34H48N8O9	$[M+2H]^{2+}$	75317	357.18380	-1.9	4.89	N/A	NA	NA
i2	Hordatine A glucoside	C34H48N8O9	$[M+2H]^{2+}$	75318	357.18350	-2.8	5.15	N/A	NA	NA
j1	Hordatine B glucoside	C35H49N8O10	$[M+2H]^{2+}$	79927	372.18920	-1.5	4.76	N/A	3.1E-01	1.21
j2	Hordatine B glucoside	C35H49N8O10	$[M+2H]^{2+}$	79928	372.18930	-1.3	5.13	N/A	3.1E-01	1.21
k1	Hordatine C glucoside	C36H52N8O11	$[M+2H]^{2+}$	84265	387.19520	0.4	4.91	N/A	NA	NA
k2	Hordatine C glucoside	C36H52N8O11	$[M+2H]^{2+}$	84266	387.19480	-0.7	5.26	14951	9.8E-01	3.71
l1	Hordatine A	C28H38N8O4	$[M+2H]^{2+}$	51157	276.15770	-1.4	5.74	1883	4.8E-01	2.84
l2	Hordatine A	C28H38N8O4	$[M+2H]^{2+}$	51148	276.15730	-2.8	6.09	N/A	NA	NA
m1	Hordatine B	C29H40N8O5	$[M+2H]^{2+}$	55514	291.16270	-2.3	5.61	1008162	9.7E-01	3.85
m2	Hordatine B	C29H40N8O5	$[M+2H]^{2+}$	55506	291.16260	-2.6	6.06	N/A	NA	NA
n	Hordatine C	C30H42N8O6	$[M+2H]^{2+}$	59788	306.16760	-3.4	6.22	N/A	6.3E-01	0.72



Supplementary figure 31: Heatmap of previously annotated secondary metabolites including cyano glycoside derivatives (components **a**–**d** and **f**, $[M–H]^-$) and hordatine derivatives (components **i**–**n**, $[M+2H]^{2+}$) in barley in different experiments. Ratios of log2-transformed fold changes are given by shade of red or blue according to the scale bar. Data represent mean values of six replicates. Statistical analysis was performed with ANOVA and post-hoc Tukey's test, of which * *p*-value <0.05, ** < 0.01 and *** < 0.001.

Supplementary table 8: Number of significant metabolites or metabolic features detected in individual batches

Expression level *Lr34*	**Primary metabolites**	Barley			Rice		
		BH	BM	BI	RH	RM	RI
low	# significant in batch 1	4	7	10	7	2	1
	# significant in batch 2	3	4	3	7	5	3
	Intersection batch 1+2	**1**	**3**	**0**	**0**	**1**	**0**
high	# significant in batch 1	27	72	70	24	1	3
	# significant in batch 2	27	65	62	44	10	10
	Intersection batch 1+2	**11**	**50**	**53**	**7**	**0**	**1**

Expression level *Lr34*	**Lipids**	Barley			Rice		
		BH	BM	BI	RH	RM	RI
low	# significant in batch 1	2	0	8	7	0	1
	# significant in batch 2	0	0	5	2	15	7
	Intersection batch 1+2	**0**	**0**	**0**	**0**	**0**	**0**
high	# significant in batch 1	38	94	77	17	1	0
	# significant in batch 2	43	86	95	13	1	8
	Intersection batch 1+2	**33**	**69**	**43**	**0**	**0**	**0**

Expression level *Lr34*	**LC-(–)-ESI-MS features secondary metabolites**	Barley			Rice		
		BH	BM	BI	RH	RM	RI
low	# significant in batch 1	451	724	573	286	451	530
	# significant in batch 2	623	417	398	389	1006	1440
	Intersection batch 1+2	**243**	**176**	**36**	**67**	**98**	**182**
high	# significant in batch 1	1129	6793	7173	2219	1611	809
	# significant in batch 2	918	5086	7059	3870	2554	1803
	Intersection batch 1+2	**392**	**4018**	**5248**	**1154**	**702**	**331**

Expression level *Lr34*	**LC-(+)-ESI-MS features secondary metabolites**	Barley			Rice		
		BH	BM	BI	RH	RM	RI
low	# significant in batch 1	60	548	495	181	392	141
	# significant in batch 2	214	309	309	96	663	389
	Intersection batch 1+2	**20**	**118**	**3**	**2**	**34**	**39**
high	# significant in batch 1	547	5874	6826	1377	921	294
	# significant in batch 2	388	4027	6960	1809	1606	710
	Intersection batch 1+2	**182**	**3347**	**5008**	**451**	**396**	**112**

Supplementary table 9: Detailed growth timing of barley and rice

Barley	Batch	
Hydroponics	BH1	BH2
Seed sterilization	14.11.14	28.11.14
Stratification	14-17.11.14	28.11.-01.12.14
Germination	17.-21.11.14	01.-05.12.14
Transfer hydroponics	21.11.14	05.12.14
Harvest	01.12.14,17:05	15.12.14, 17:20

Mock and pathogen infected	BM1 + BI1	BM2 + BI2
Planting	22.12.14	19.12.14
Infection	06.01.15	05.01.15
Harvest	14.01.2015, 15:50	16.01.15, 11:10

Rice	Batch	
Hydroponics	RH1	RH2
Seed de-husking	12.12.14	09.01.15
Transfer hydroponics	19.12.14	16.01.15
Harvest	07.01.15, 10:05	04.02.15, 11:20

Mock and pathogen infected	RM1 + RI1	RM2 + RI2
Planting	05.01.15	13.01.15
Infection	27.01.15, 10:00	04.02.15, 07:30
Harvest	28.01.15, 14:00	05.02.15, 12:15

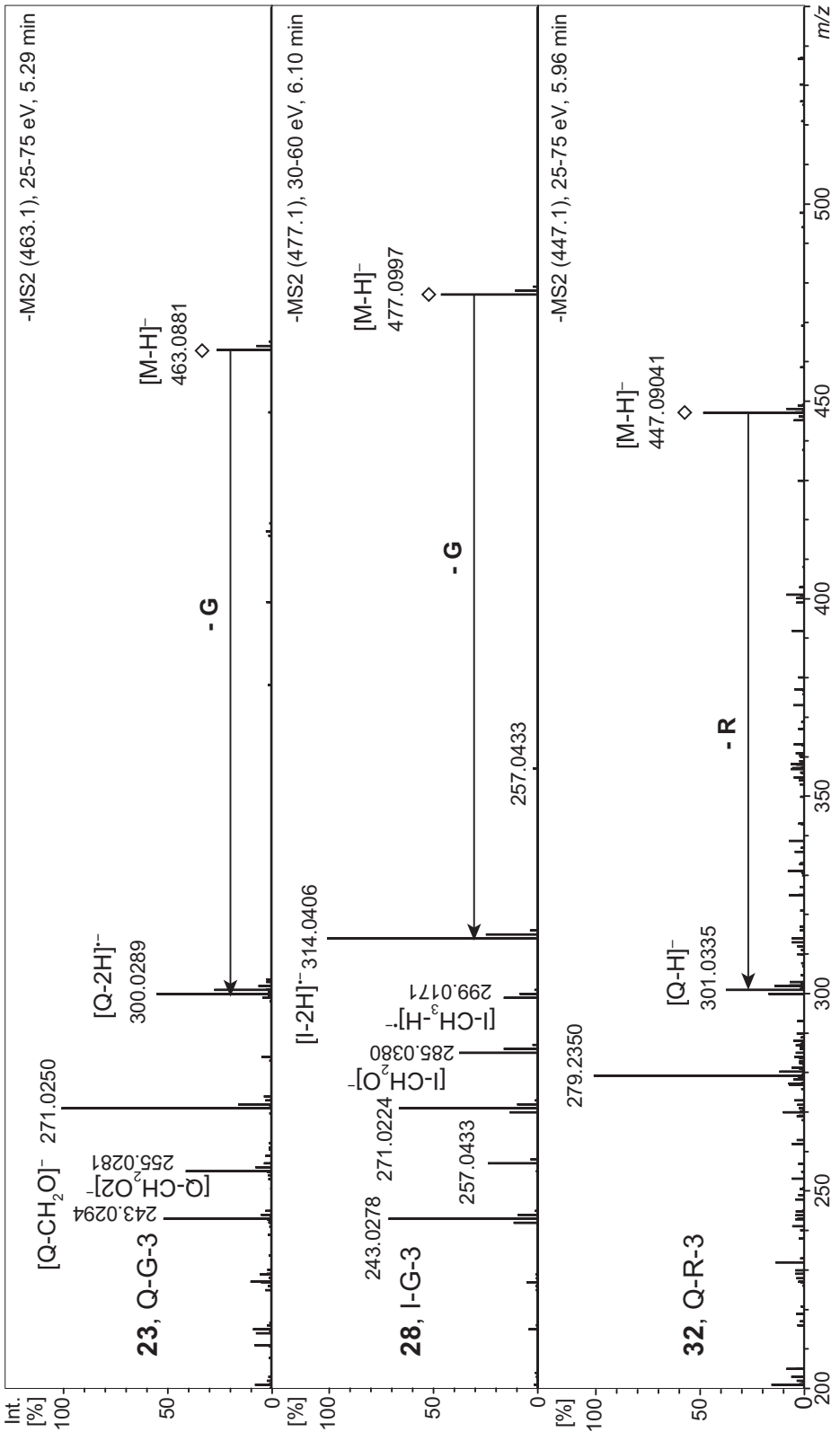
Supplementary table 10: Quality control mixture for GC-MS analysis

Name	Formula	CAS	Weight [mg]
4-Hydroxybenzoic acid	$C_7H_6O_3$	99-96-7	10.0
α-Ketoglutaric acid	$C_5H_6O_5$	328-50-7	10.1
Alanine	$C_3H_7NO_2$	338-69-2	9.9
Caffeic acid	$C_9H_8O_4$	331-39-5	10.4
Cholesterol	$C_{27}H_{46}O$	57-88-5	10.0
Citramalic acid	$C_5H_8O_5$	08-10-36	5.0
Citric acid	$C_6H_8O_7$	77-92-9	10.1
Fucose	$C_6H_{12}O_5$	2438-80-4	9.7
Glucose	$C_6H_{12}O_6$	492-62-6	10.3
Glycine	$C_2H_5NO_2$	56-40-6	10.0
Isoleucine	$C_6H_{13}NO_2$	443-79-8	10.0
Lactitol	$C_{12}H_{26}O_{12}$	81025-04-9	5.1
Lactose	$C_{12}H_{22}O_{11}$	5965-66-2	5.0
Lanosterol	$C_{30}H_{50}O$	79-63-0	10.3
Maltose	$C_{12}H_{24}O_{12}$	6363-53-7	20.0
Maltotrieose	$C_{18}H_{32}O_{16}$	1109-28-0	10.2
Palatinose	$C_{12}H_{22}O_{11}$	13718-94-0	10.0
Calcium pantothenate	$C_{18}H_{32}CaN_2O_{10}$	137-08-6	10.3
Putrescine	$C_4H_{14}Cl_2N_2$	333-93-7	9.8
3-Hydroxypyridine	C_5H_5NO	109-00-2	9.8
Ribitol	$C_5H_{12}O_5$	488-81-3	10.2
Ribose	$C_5H_{10}O_5$	50-69-1	10.0
Sorbitol	$C_6H_{14}O_6$	50-70-4	5.0
Sorbose	$C_6H_{12}O_6$	87-79-6	10.0
Stigmasterol	$C_{29}H_{48}O$	83-48-7	9.9
Threitol	$C_4H_{10}O_4$	6968-16-7	9.9
Urea	CH_4N_2O	57-13-6	9.8
Valin	$C_5H_{11}NO_2$	72-18-4	10.2

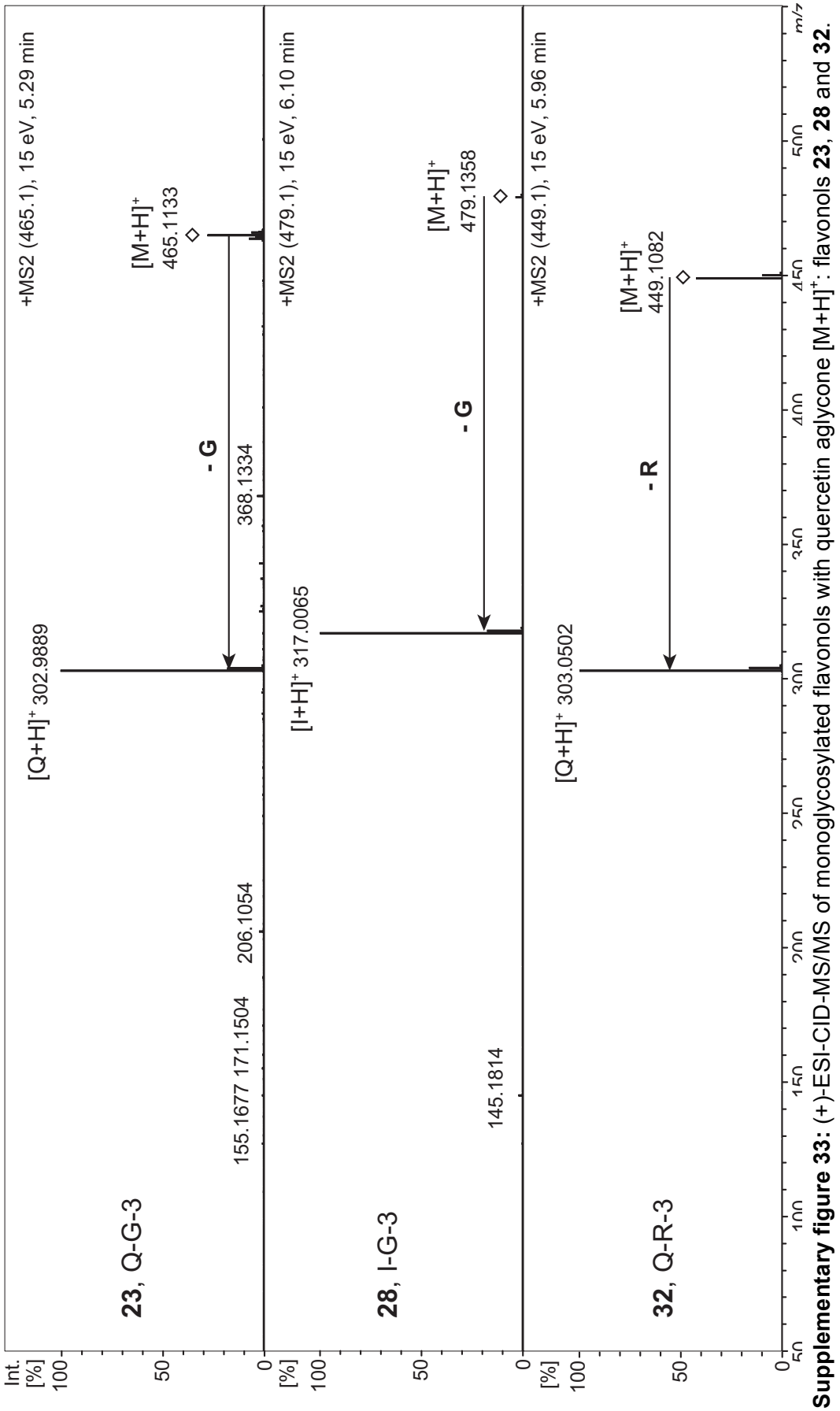
Supplementary table 11: FAME mixture for retention time indices of GC-MS analyses

Name	Formula	C-chain	CAS-No.	Weight [mg]
Methylcaprylate	$C_9H_{18}O_2$	C8	111-11-5	20
Methylperlargonate	$C_{10}H_{20}O_2$	C9	1731-84-6	20
Methylcaprate	$C_{11}H_{22}O_2$	C10	110-42-9	20
Methyllaurate	$C_{13}H_{26}O_2$	C12	111-82-0	20
Methylmyristate	$C_{15}H_{30}O_2$	C14	124-10-7	20
Methylpalmitate	$C_{17}H_{34}O_2$	C16	112-39-0	20
Methylstearate	$C_{19}H_{38}O_2$	C18	112-61-8	10
Methyleicosanoate	$C_{21}H_{42}O_2$	C20	1120-28-1	10
Methyldocosanoate	$C_{23}H_{46}O_2$	C22	929-77-1	10
Lignoceric acid methylester	$C_{25}H_{50}O_2$	C24	2442-49-1	10
Methylhexacosanoate	$C_{27}H_{54}O_2$	C26	5802-82-4	10
Methyloctacosanoate	$C_{29}H_{58}O_2$	C28	55682-92-3	10
Triacontanoic acid methylester	$C_{31}H_{62}O_2$	C30	629-83-4	10

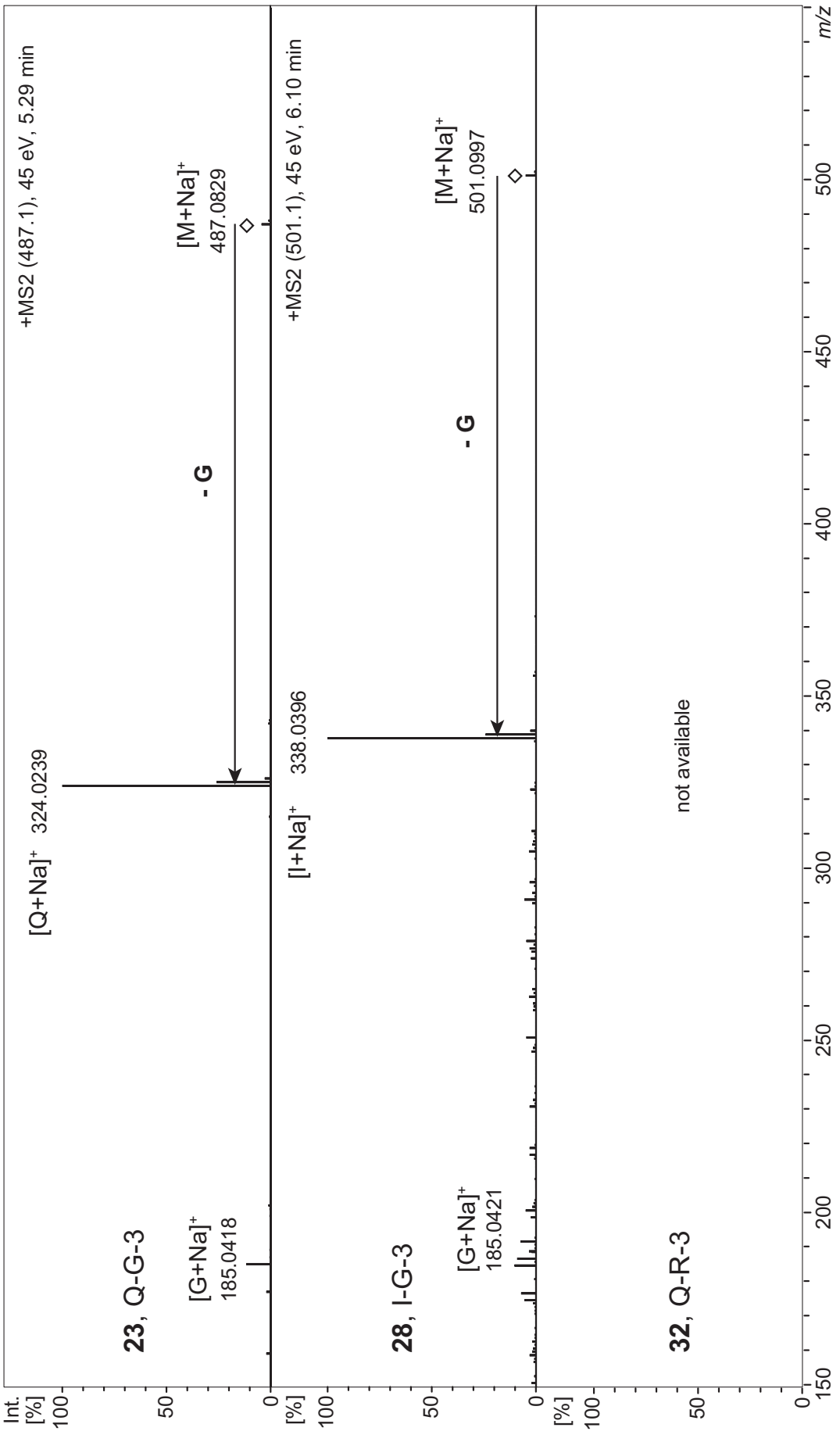
To chapter 3: Targeted analysis of flavonols in *Arabidopsis thaliana*



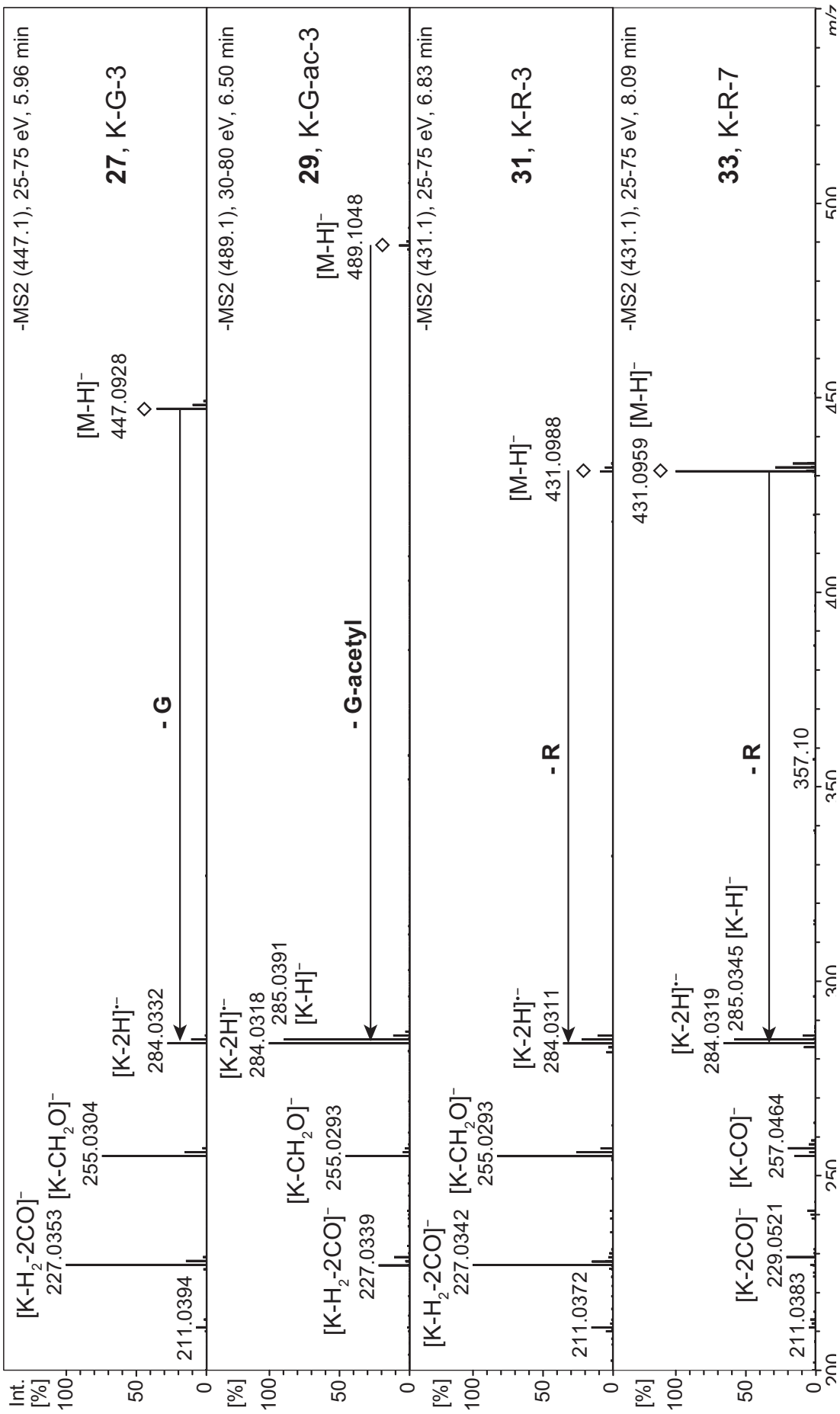
Supplementary figure 32: (–)-ESI-CID-MS/MS of monoglycosylated flavonols with quercetin aglycone $[M–H]^-$: flavonols **23**, **28** and **32**.



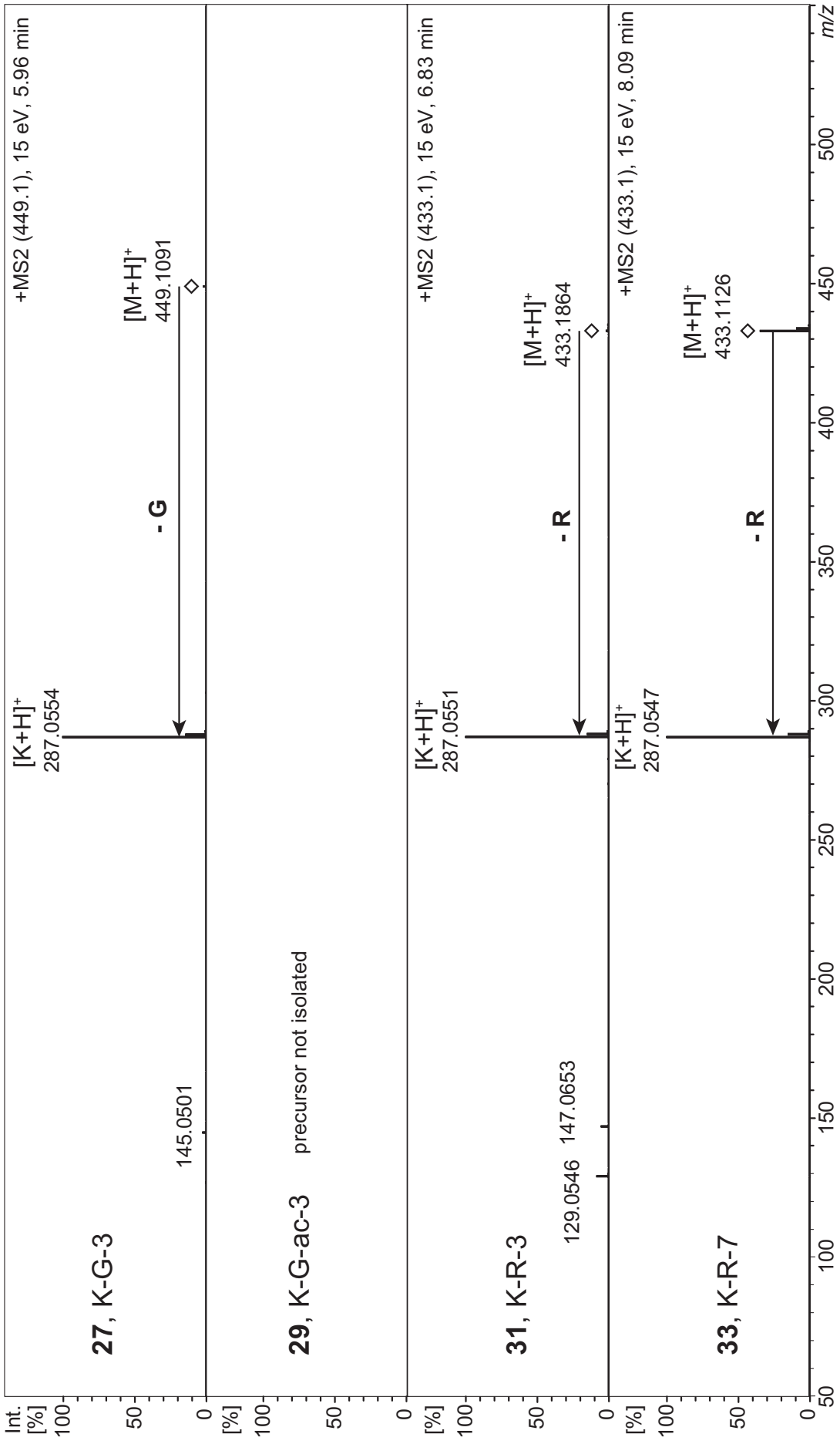
Supplementary figure 33: (+)-ESI-CID-MS/MS of monoglycosylated flavonols with quercetin aglycone $[M+H]^+$: flavonols **23**, **28** and **32**.



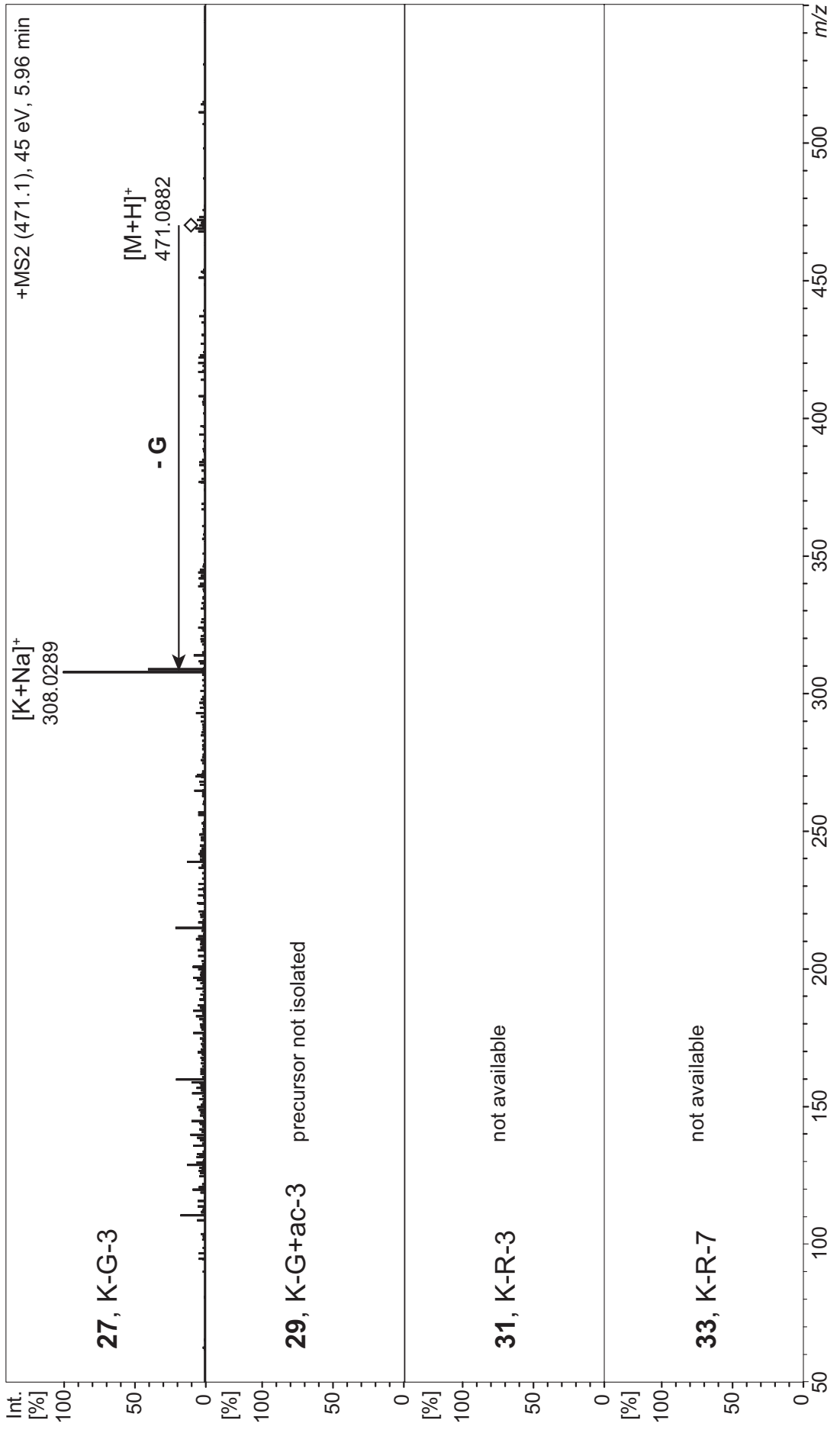
Supplementary figure 34: (+)-ESI-CID-MS/MS of monoglycosylated flavonols with quercetin aglycone $[M+Na]^{+\cdot}$: flavonols **23**, and **28**.



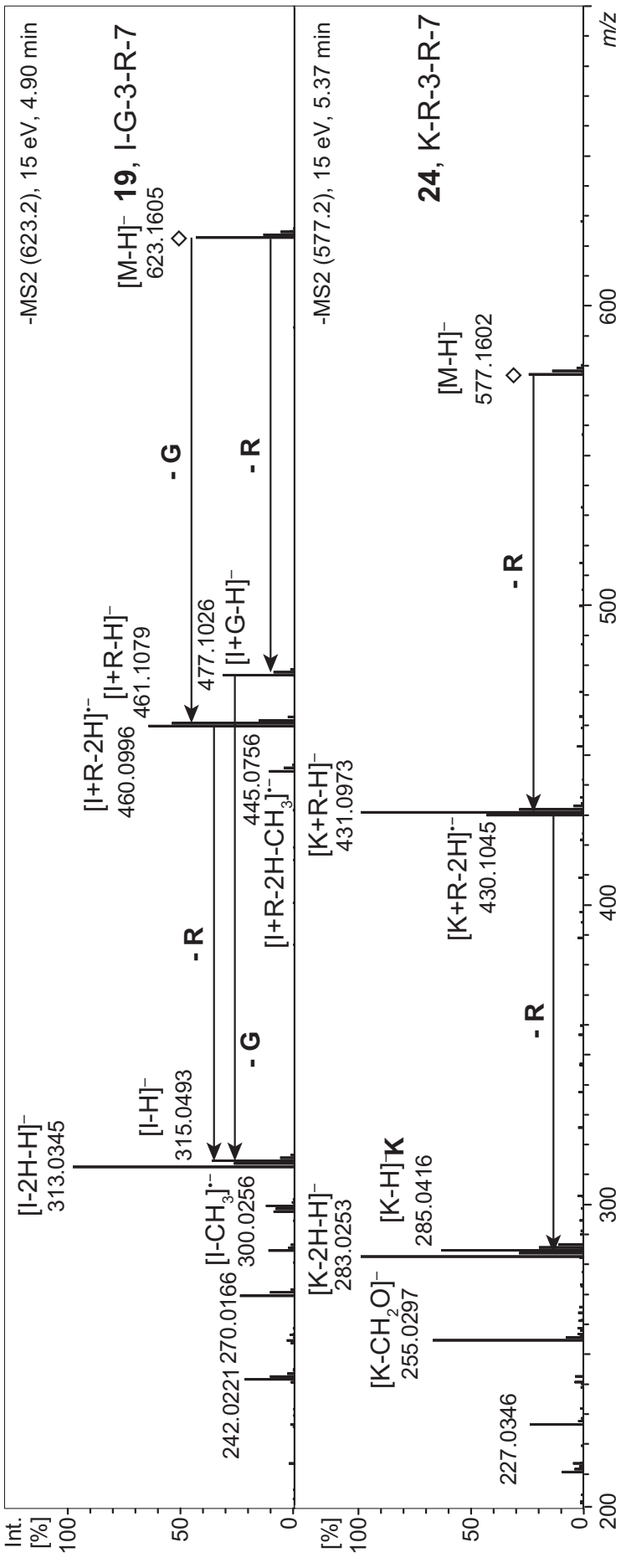
Supplementary figure 35: (–)-ESI-CID-MS/MS of monoglycosylated flavonols with kaempferol aglycone $[M–H]^-$: flavonols **27**, **29**, **31** and **33**.



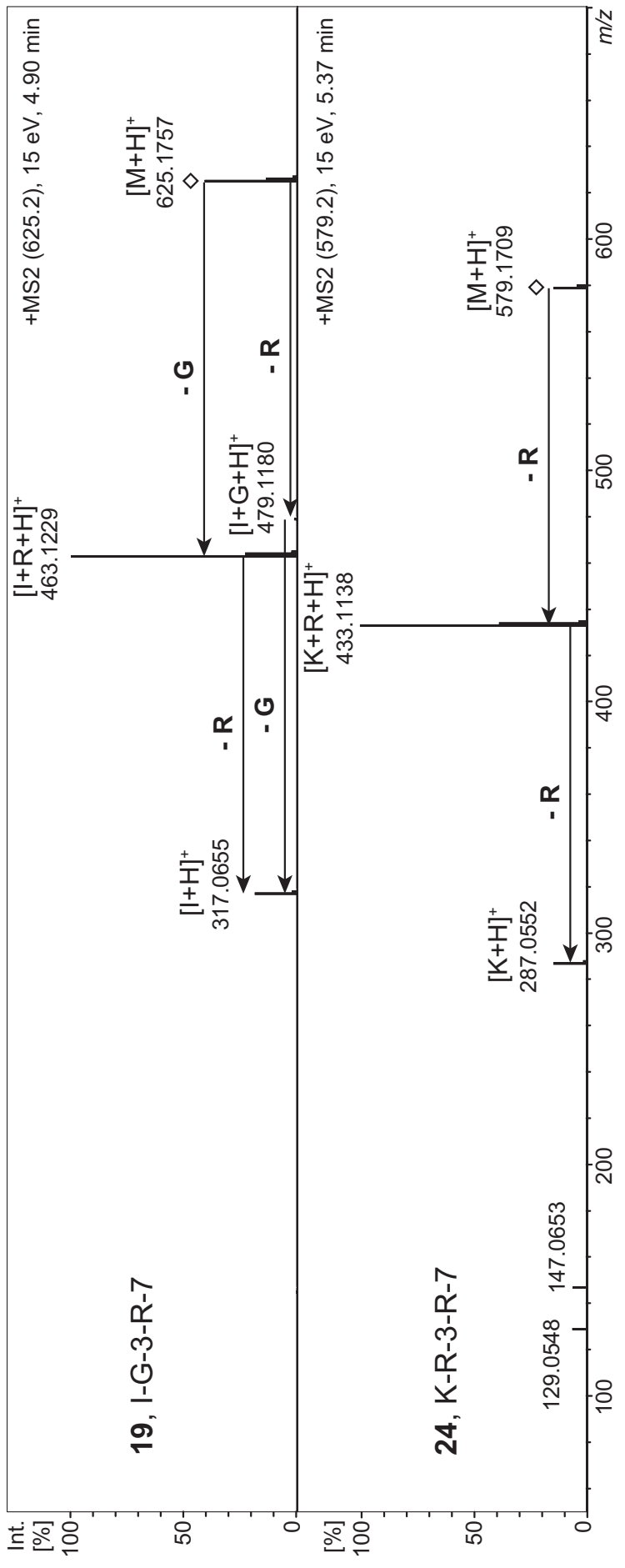
Supplementary figure 36: (+)-ESI-CID-MS/MS of monoglycosylated flavonols with kaempferol aglycone $[M+H]^+$: flavonols **27**, **31** and **33**.



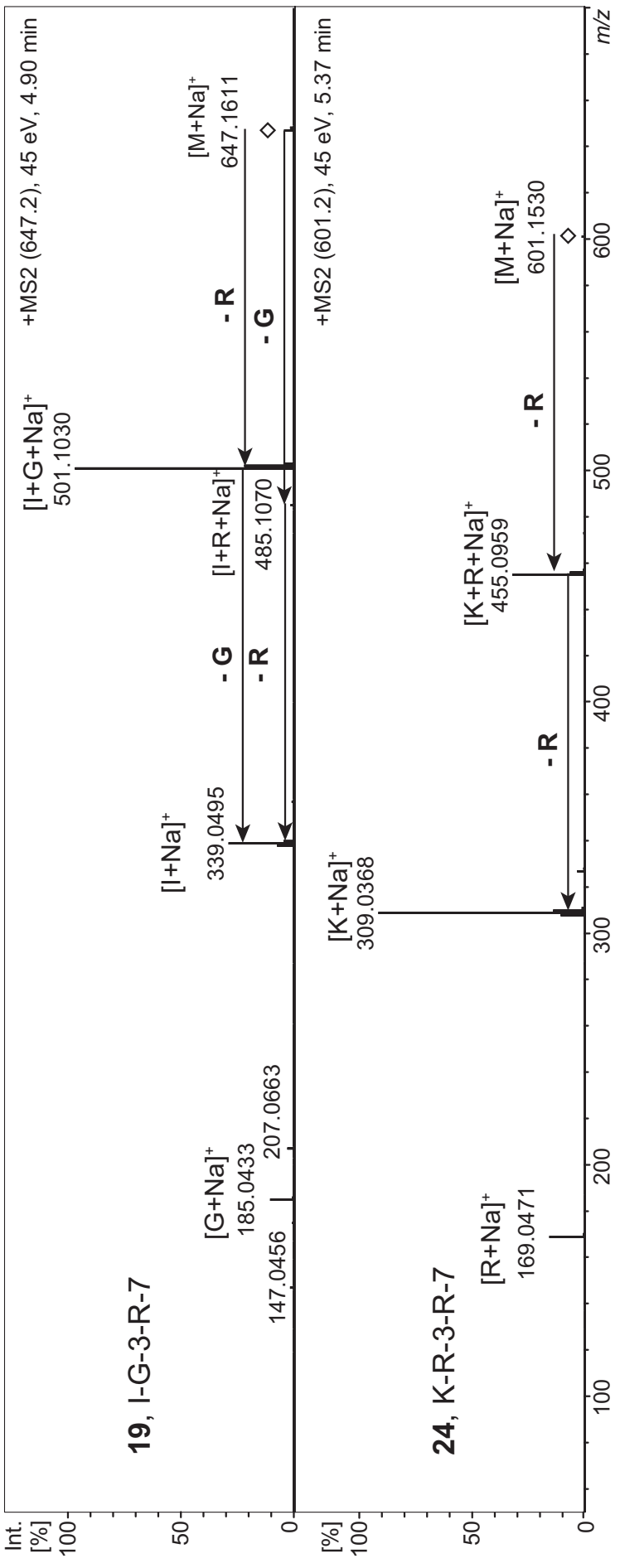
Supplementary figure 37: (+)-ESI-CID-MS/MS of monoglycosylated flavonols with kaempferol aglycone $[M+Na]^+$: flavonol **27**.



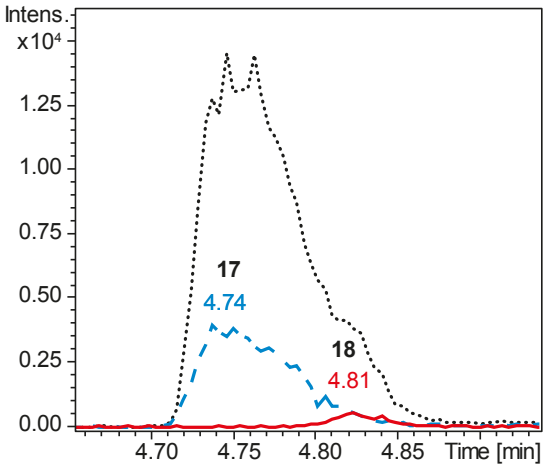
Supplementary figure 38: (–)-ESI-CID-MS/MS of diglycosylated flavonols $C_{28}H_{32}O_{16}$ and $C_{21}H_{20}O_{12}$ $[M–H]^-$: flavonols **19** and **24.**



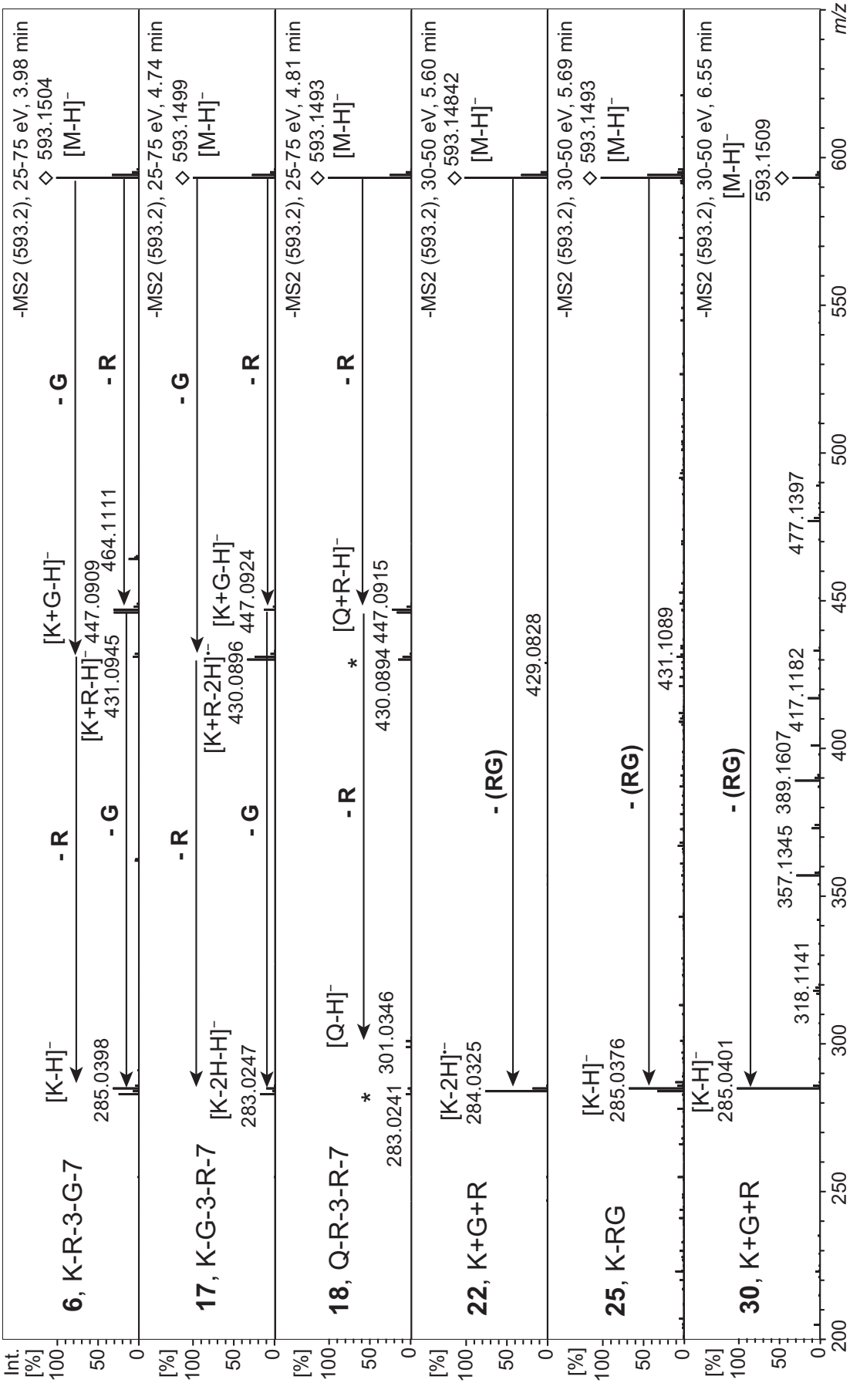
Supplementary figure 39: (+)-ESI-CID-MS/MS of diglycosylated flavonols $C_{28}H_{32}O_{16}$ and $C_{21}H_{20}O_{12}$ $[M+H]^+$: flavonols **19** and **24.**



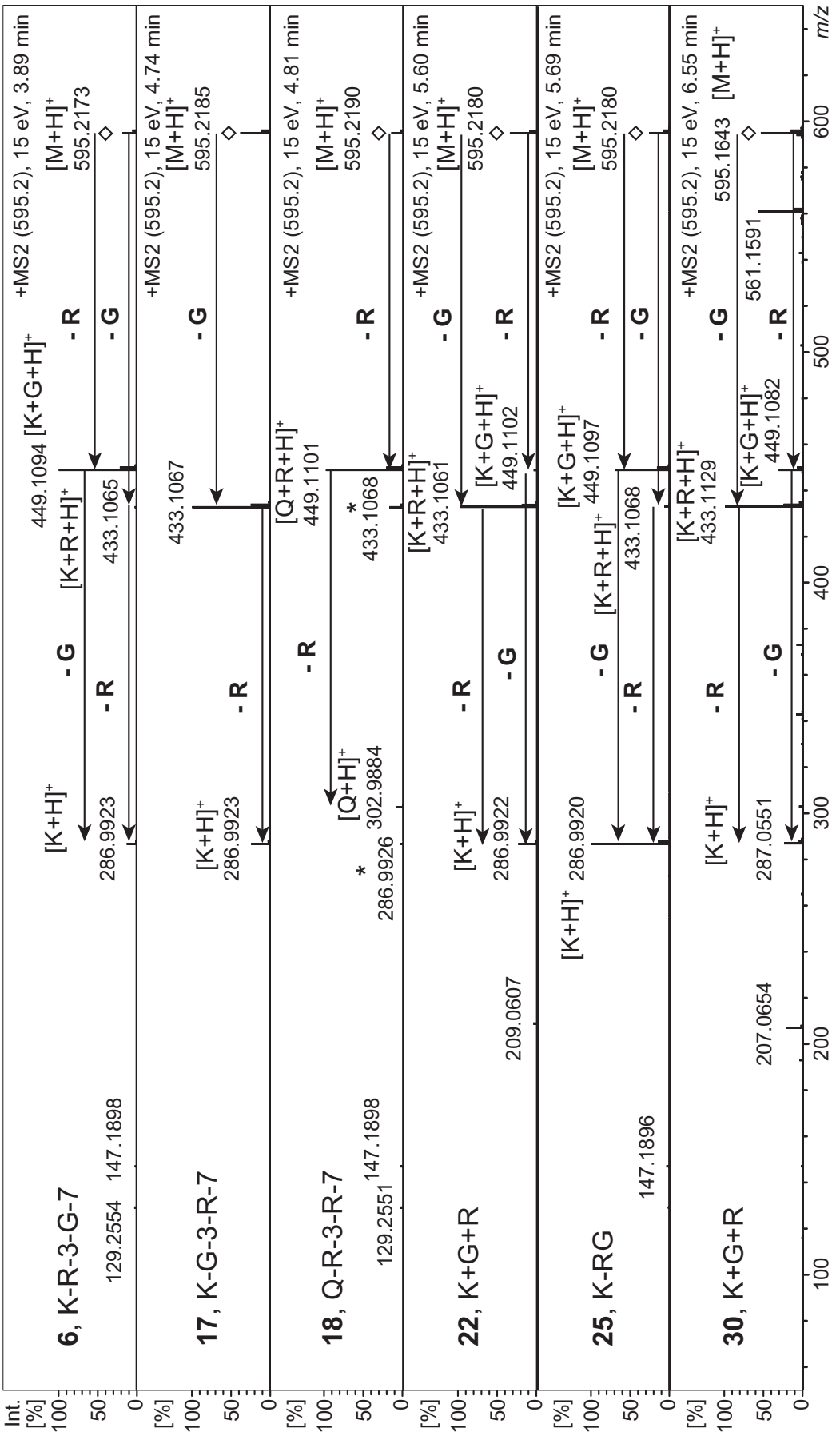
Supplementary figure 40: (+)-ESI-CID-MS/MS of diglycosylated flavonols $C_{28}H_{32}O_{16}$ and $C_{21}H_{20}O_{12}$ $[M+Na]^+$: flavonols **19** and **24.**



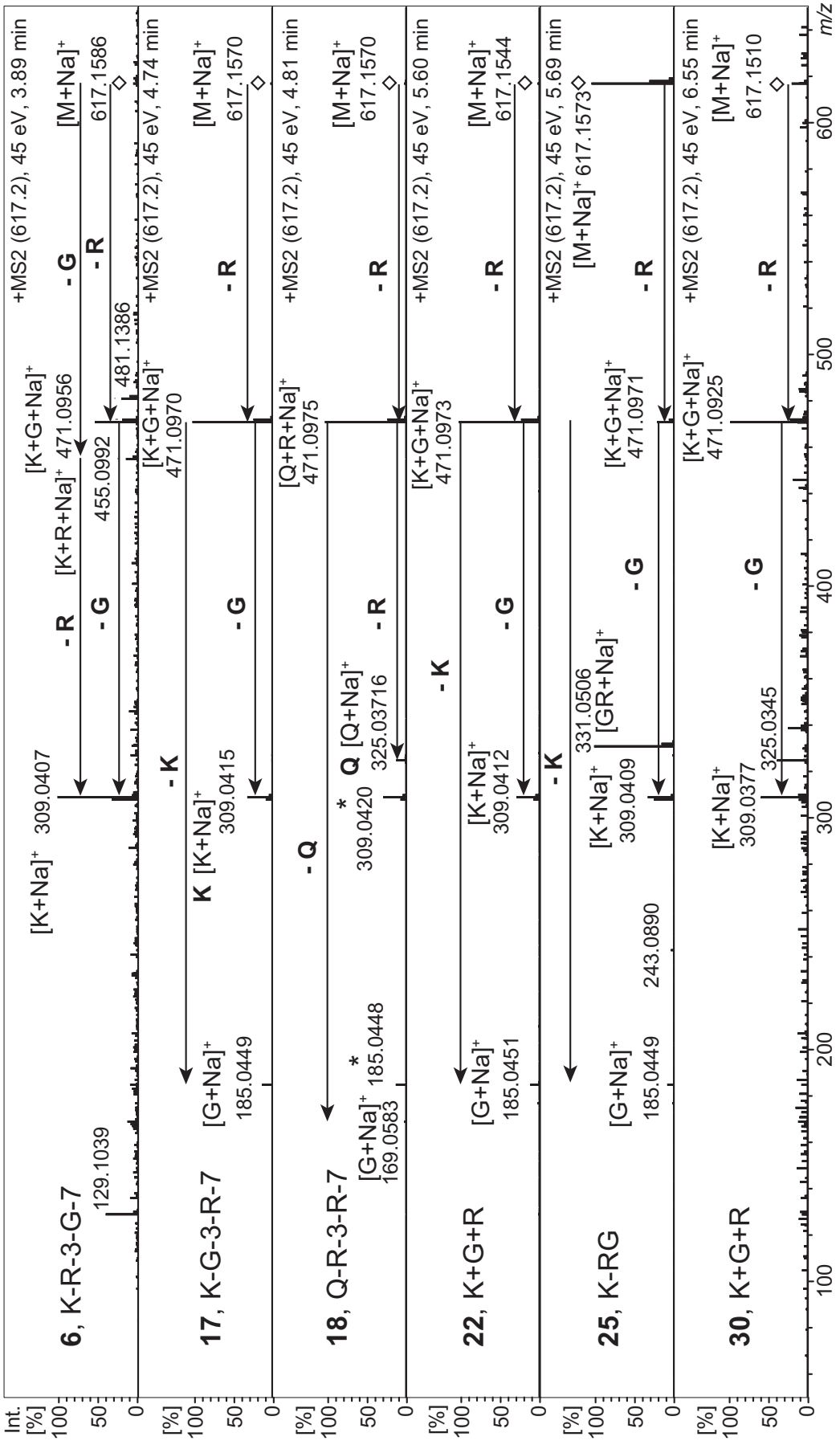
Supplementary figure 41: Extracted ion chromatograms of (–)-ESI MS^2 of $[M–H]^-$ at *m/z* 593.15 (dotted, black) and aglycone fragments $[Y_0]^-$ and $[Y_0–2H]^-$ of kaempferol **17** at *m/z* 285.04/283.02 (dashed, blue) and quercetin **18** at *m/z* 301.03/299.01 (strait, red).



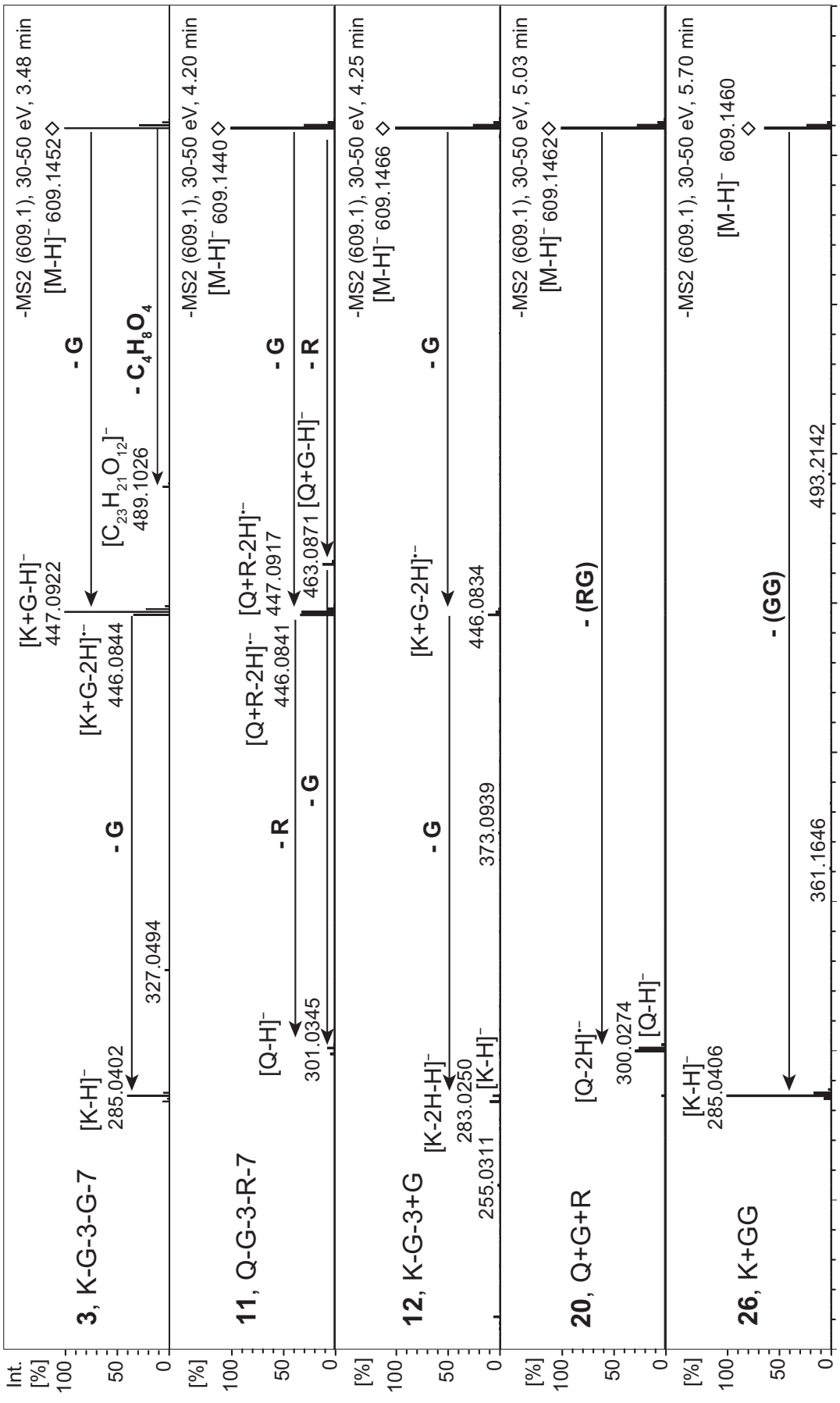
Supplementary figure 42: (–)-ESI-CID-MS/MS of diglycosylated flavonols $C_{27}H_{30}O_{15}$ $[M–H]^-$ at *m/z* 593: flavonols **6**, **17**, **18**, **22**, **25** and **30**. In spectrum of flavonol **18**, asterisks indicate overlapping fragments from flavonol **17**.



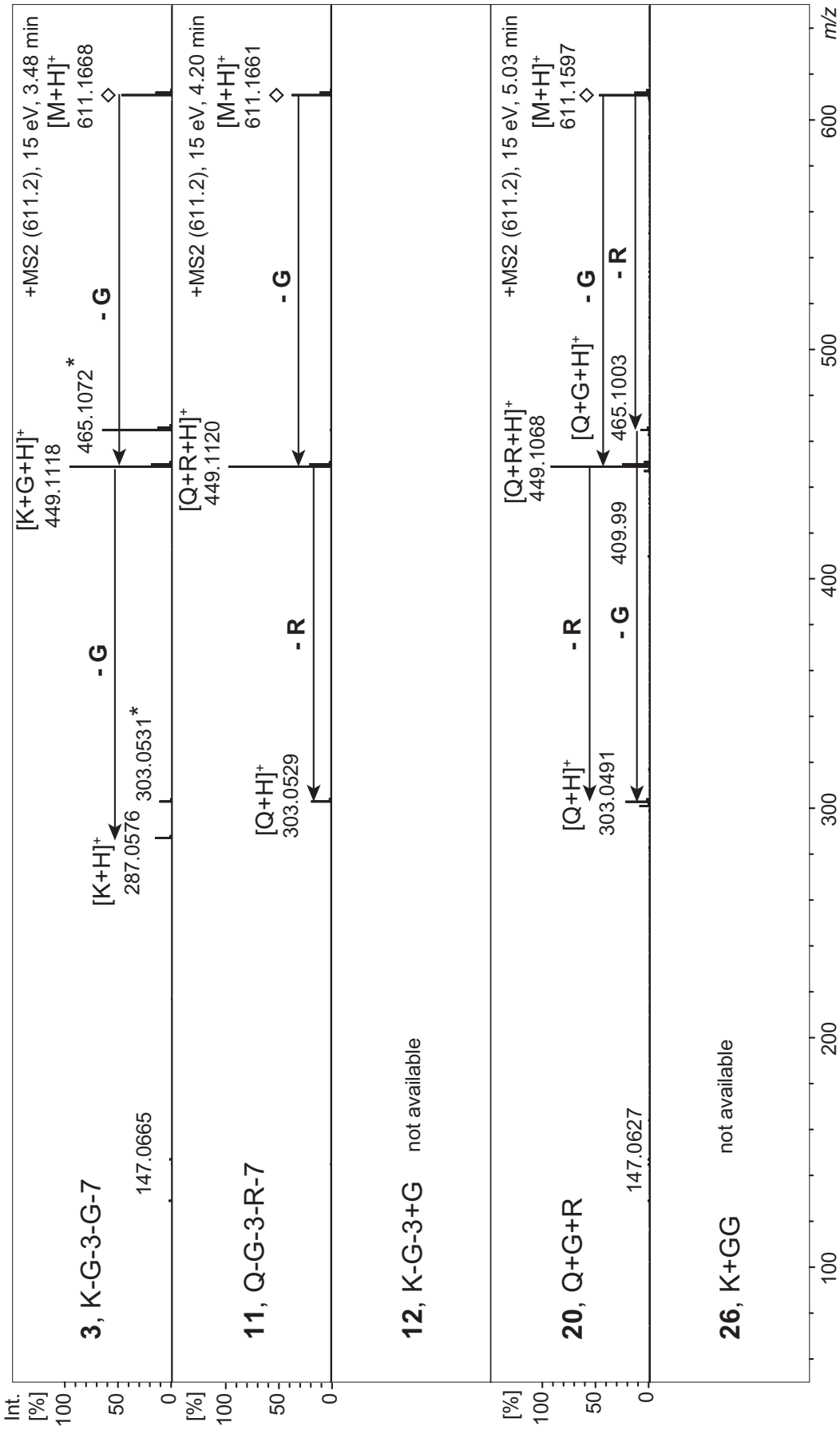
Supplementary figure 43: (+)-ESI-CID-MS/MS of diglycosylated flavonols $C_{27}H_{30}O_{15}$ $[M+H]^+$ at *m/z* 595: flavonols **6**, **17**, **18**, **22**, **25** and **30**. In spectrum of flavonol **18**, asterisks indicate overlapping fragments from flavonol **17**.



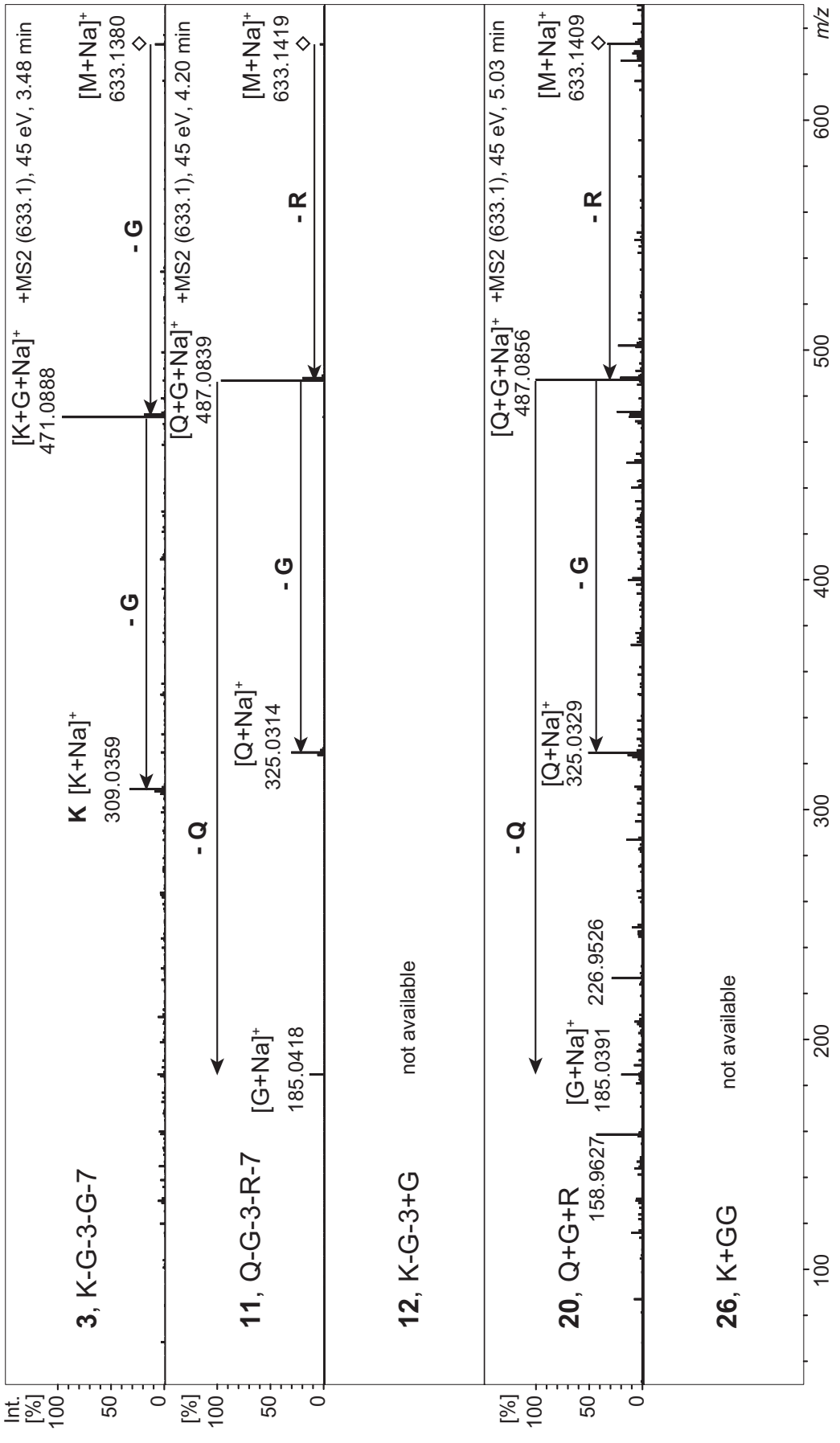
Supplementary figure 44: (+)-ESI-CID-MS/MS of diglycosylated flavonols $[M+Na]^+$ $C_{27}H_{30}O_{15}$ at *m/z* 595: flavonols **6**, **17**, **18**, **22**, **25** and **30**. In spectrum of flavonol **18**, asterisks indicate overlapping fragments from flavonol **17**.



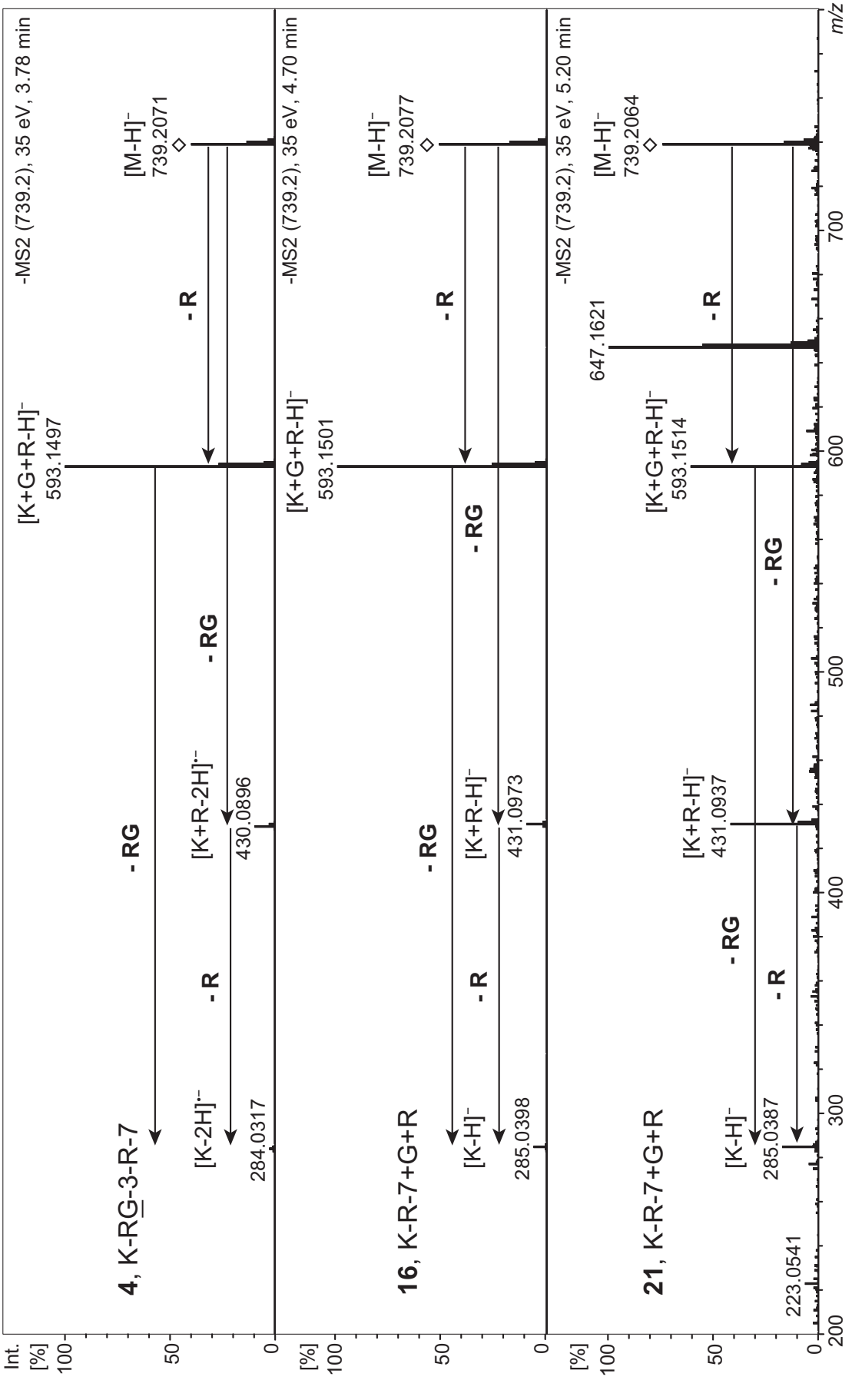
Supplementary figure 45: (–)-ESI-CID-MS/MS of diglycosylated flavonols $C_{27}H_{30}O_{16}$ $[M–H]^-$ at *m/z* 609: flavonols **3**, **11**, **12**, **20** and **26**.



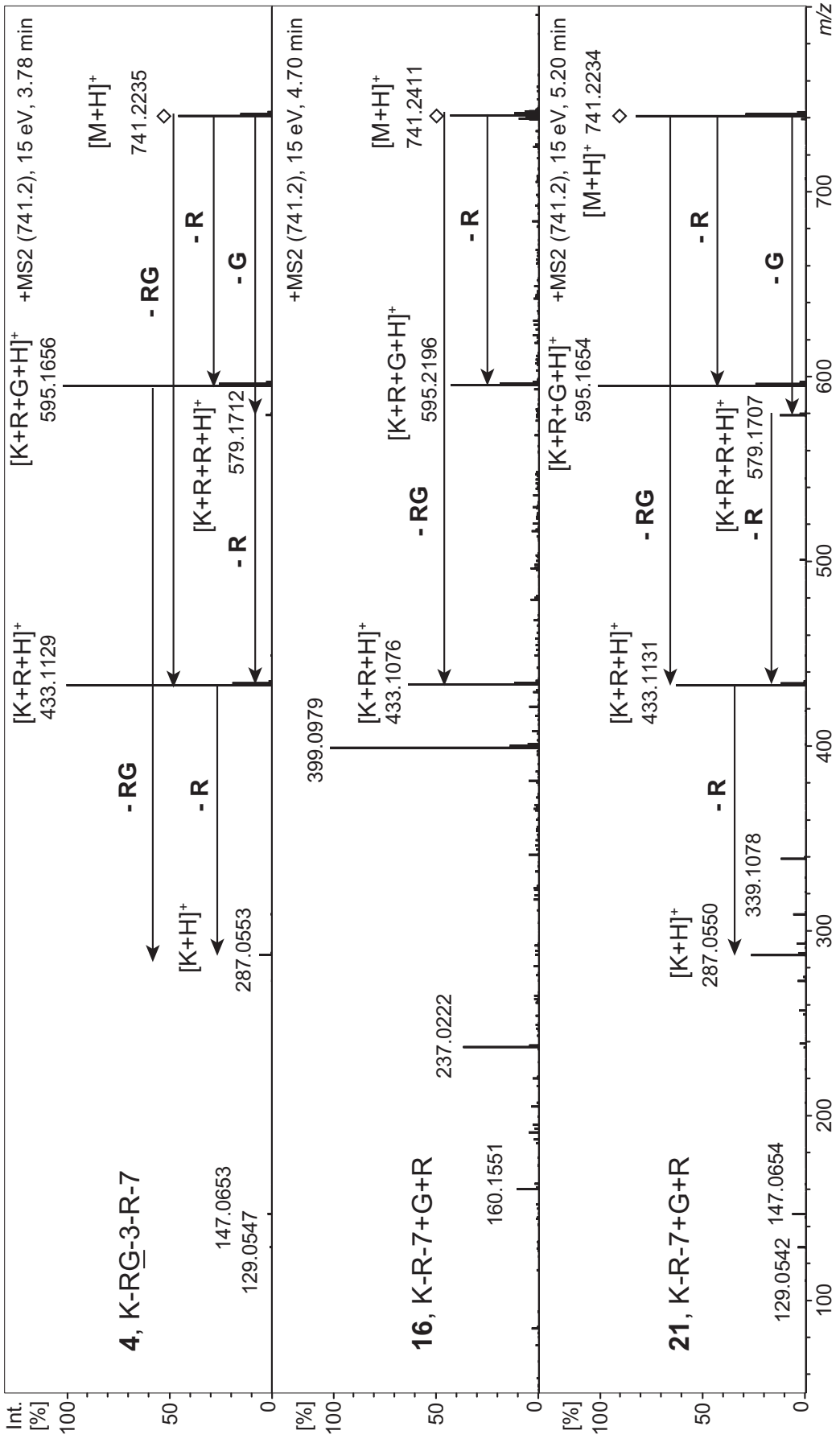
Supplementary figure 46: (+)-ESI-CID-MS/MS of diglycosylated flavonols $C_{27}H_{30}O_{16}$ $[M+H]^+$ at *m/z* 609: flavonols **3**, **11** and **20**, asterisks indicate overlapping fragments from a coeluting flavonol.



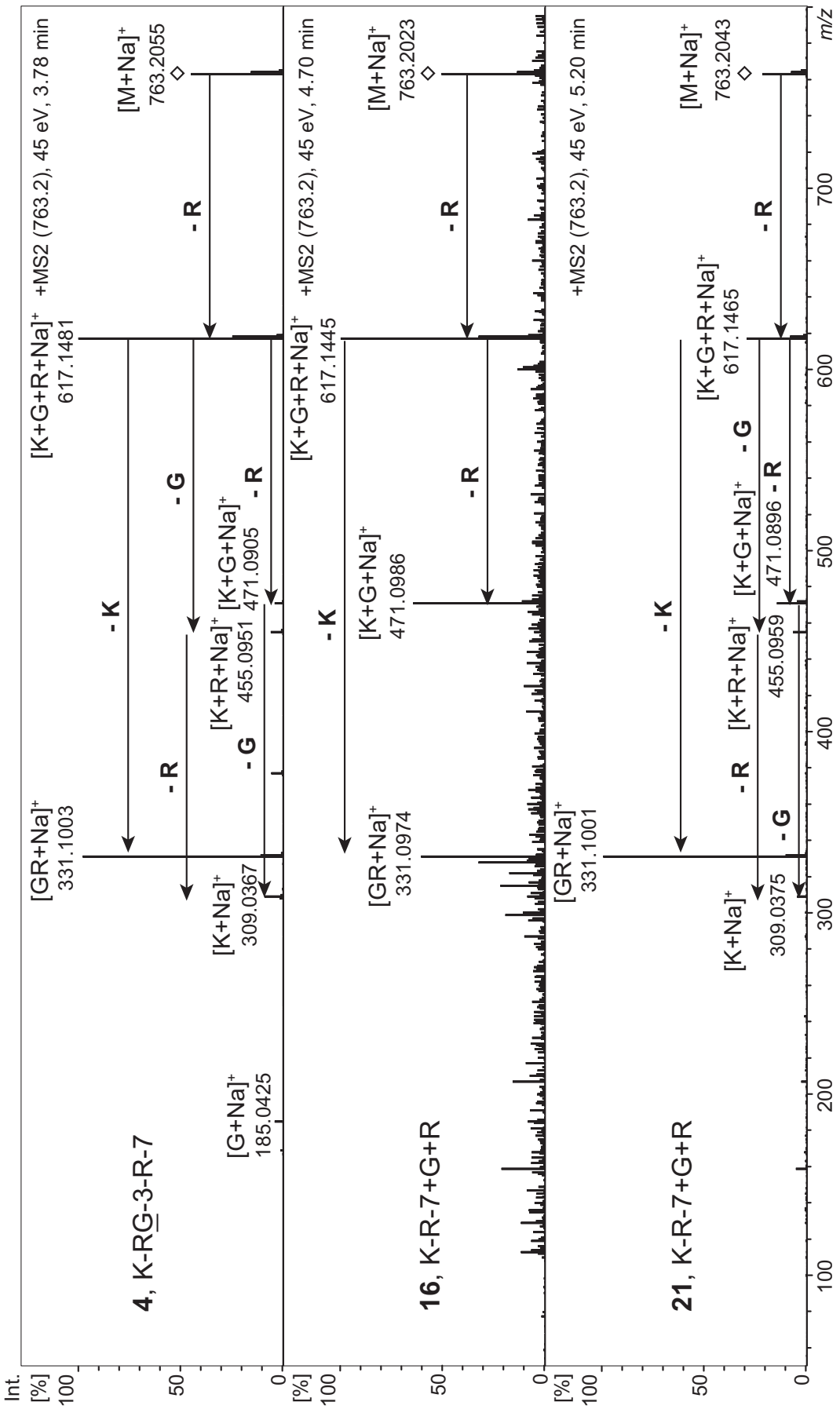
Supplementary figure 47: (+)-ESI-CID-MS/MS of diglycosylated flavonols $C_{27}H_{30}O_{16}$ $[M+Na]^+$ at *m/z* 609: flavonols **3**, **11** and **20**.



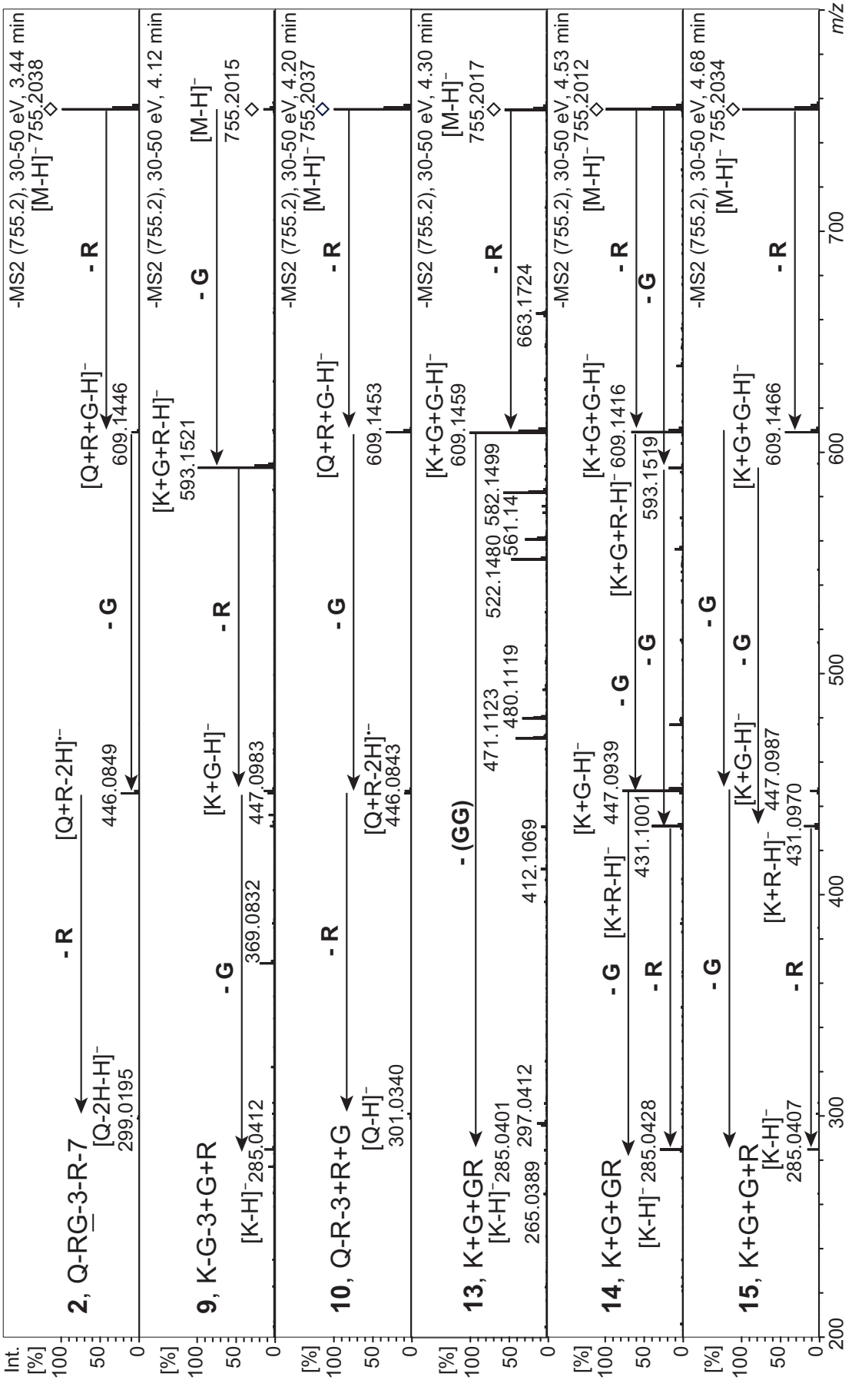
Supplementary figure 48: (–)-ESI-CID-MS/MS of diglycosylated flavonols $C_{33}H_{40}O_{19}$ $[M–H]^-$ at *m/z* 739: flavonols **4**, **16** and **21**.



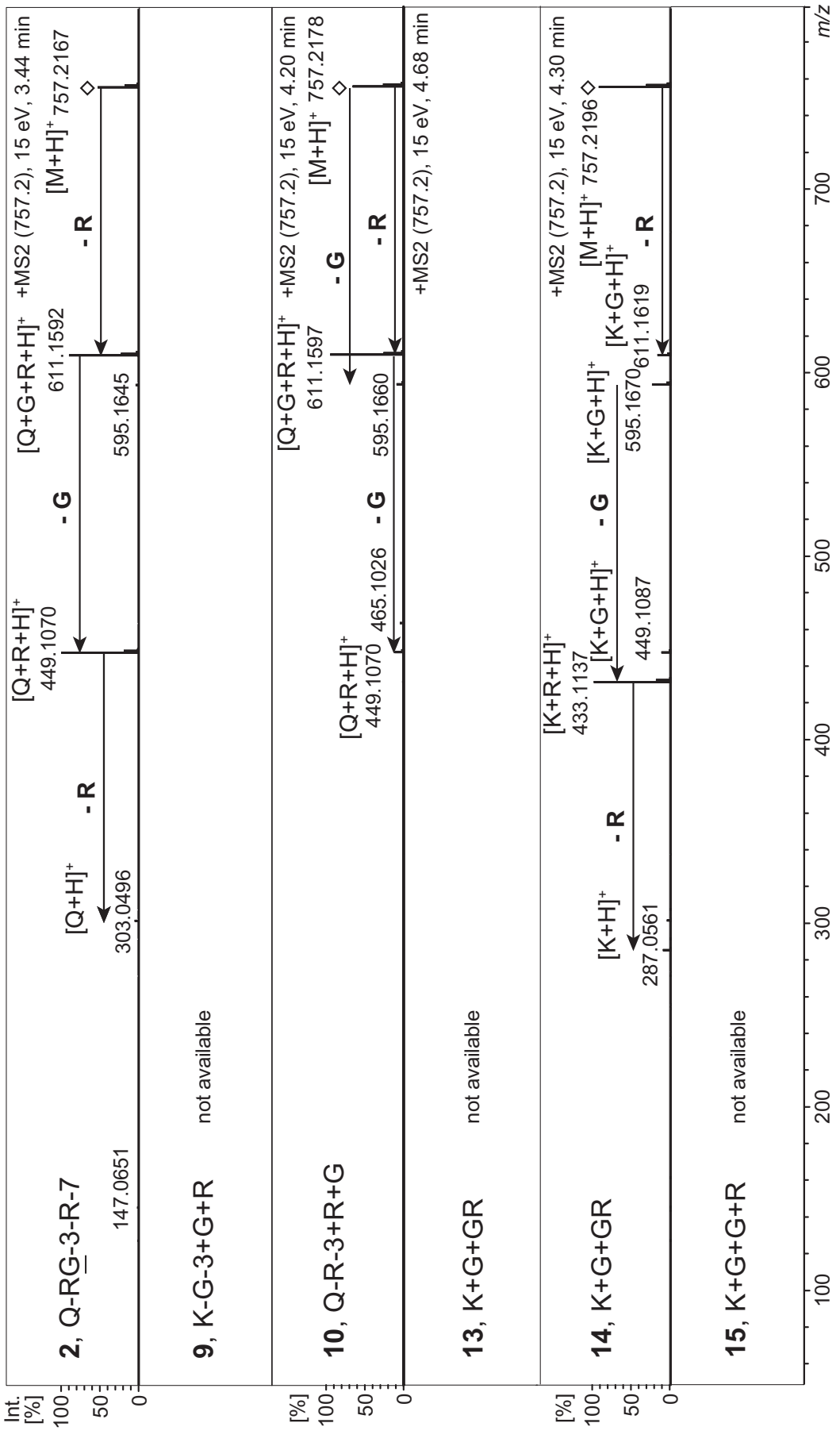
Supplementary figure 49: (+)-ESI-CID-MS/MS of diglycosylated flavonols $C_{33}H_{40}O_{19}$ $[M+H]^+$ at *m/z* 741: flavonols **4**, **16** and **21**.



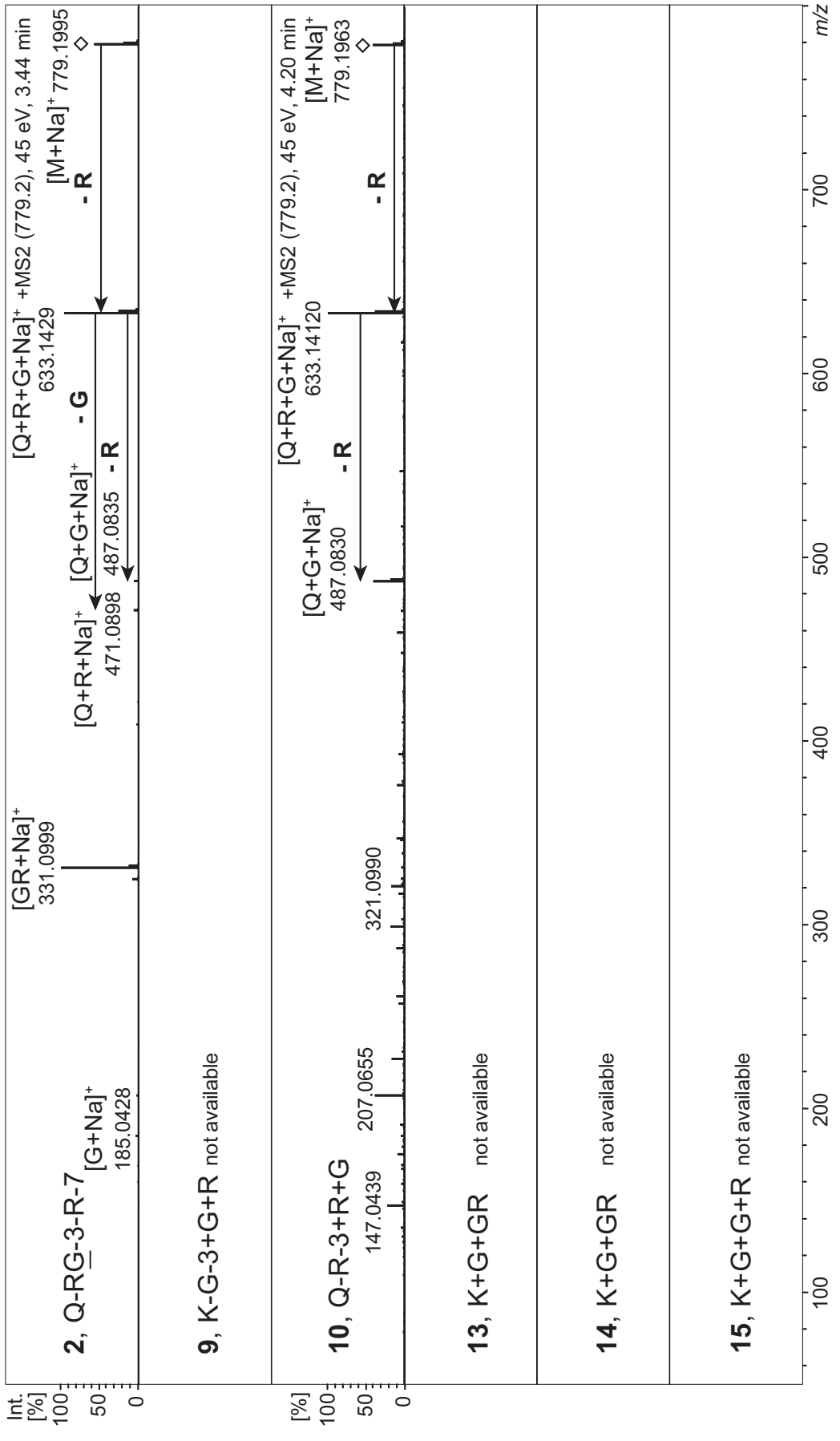
Supplementary figure 50: (+)-ESI-CID-MS/MS of triglycosylated flavonols $C_{33}H_{40}O_{19}$ $[M+Na]^+$ at *m/z* 763: flavonols **4**, **16** and **21**.



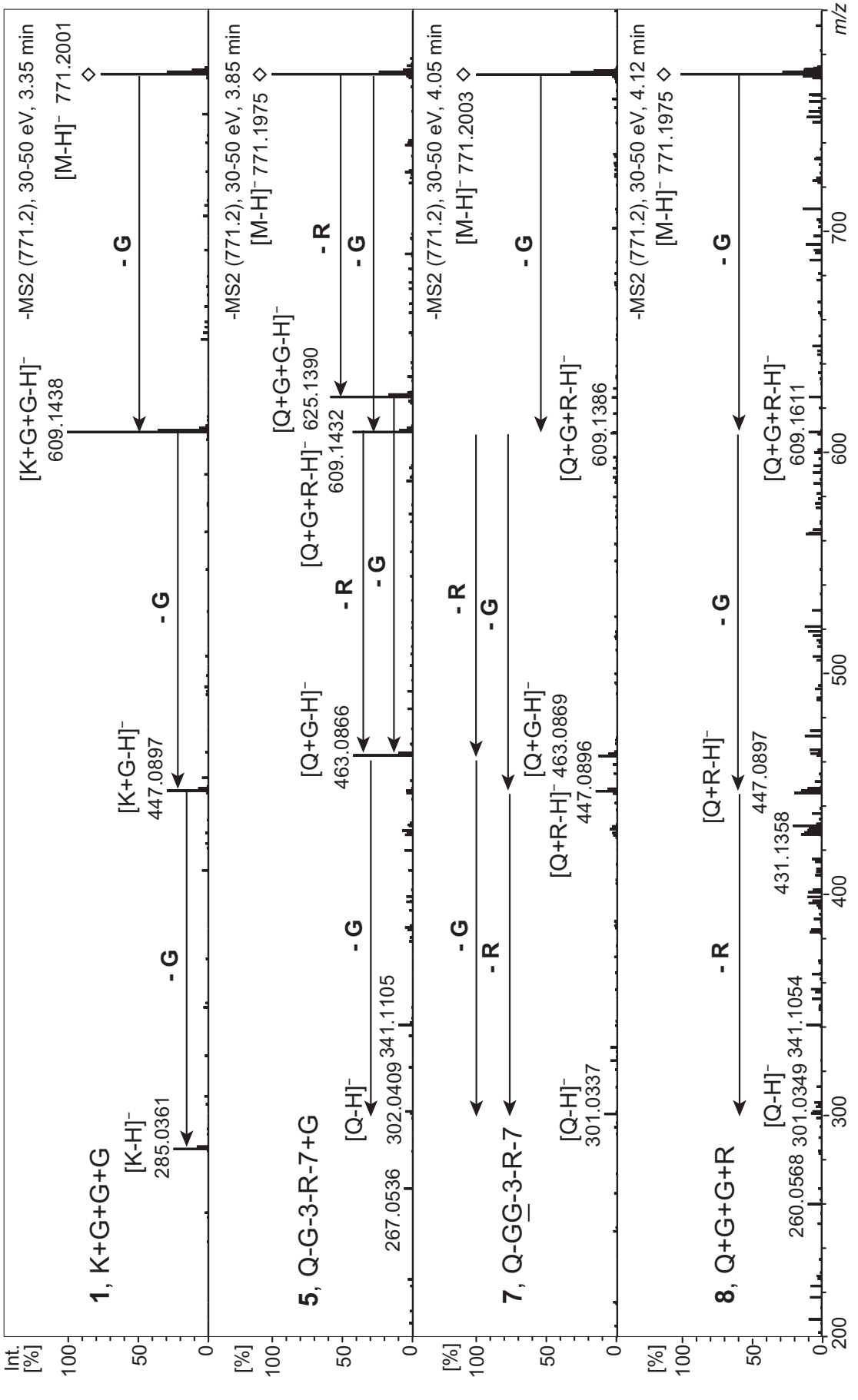
Supplementary figure 51: (–)-ESI-CID-MS/MS of triglycosylated flavonols $C_{33}H_{40}O_{20}$ [M–H]⁻ at *m/z* 755: flavonols **2**, **9**, **10**, **13**, **14**, **15**.



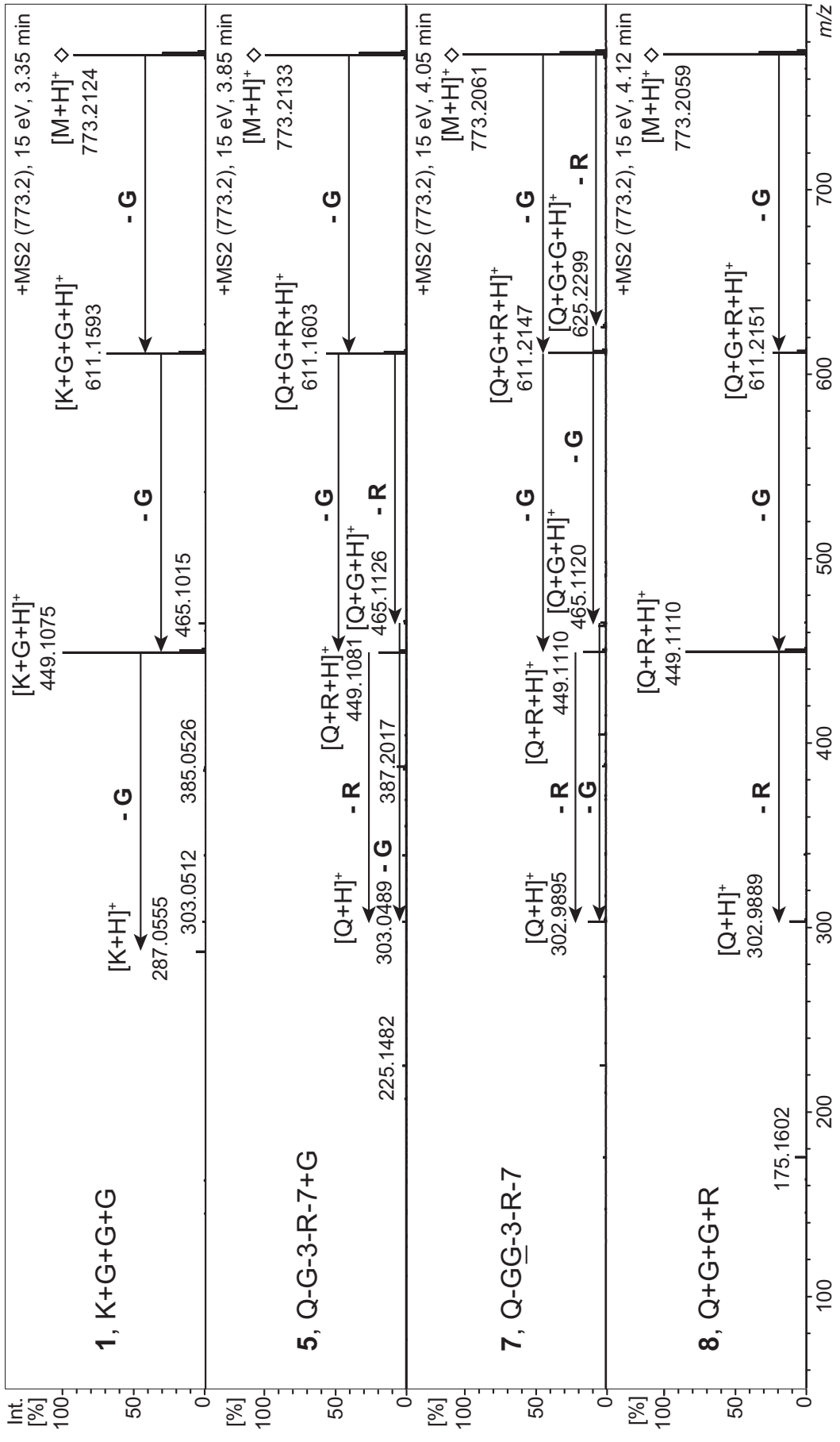
Supplementary figure 52: (+)-ESI-CID-MS/MS of triglycosylated flavonols $C_{33}H_{40}O_{20}$ $[M+H]^+$ at *m/z* 757: flavonols **2**, **9**, **10**, **13**, **14**, **15**.



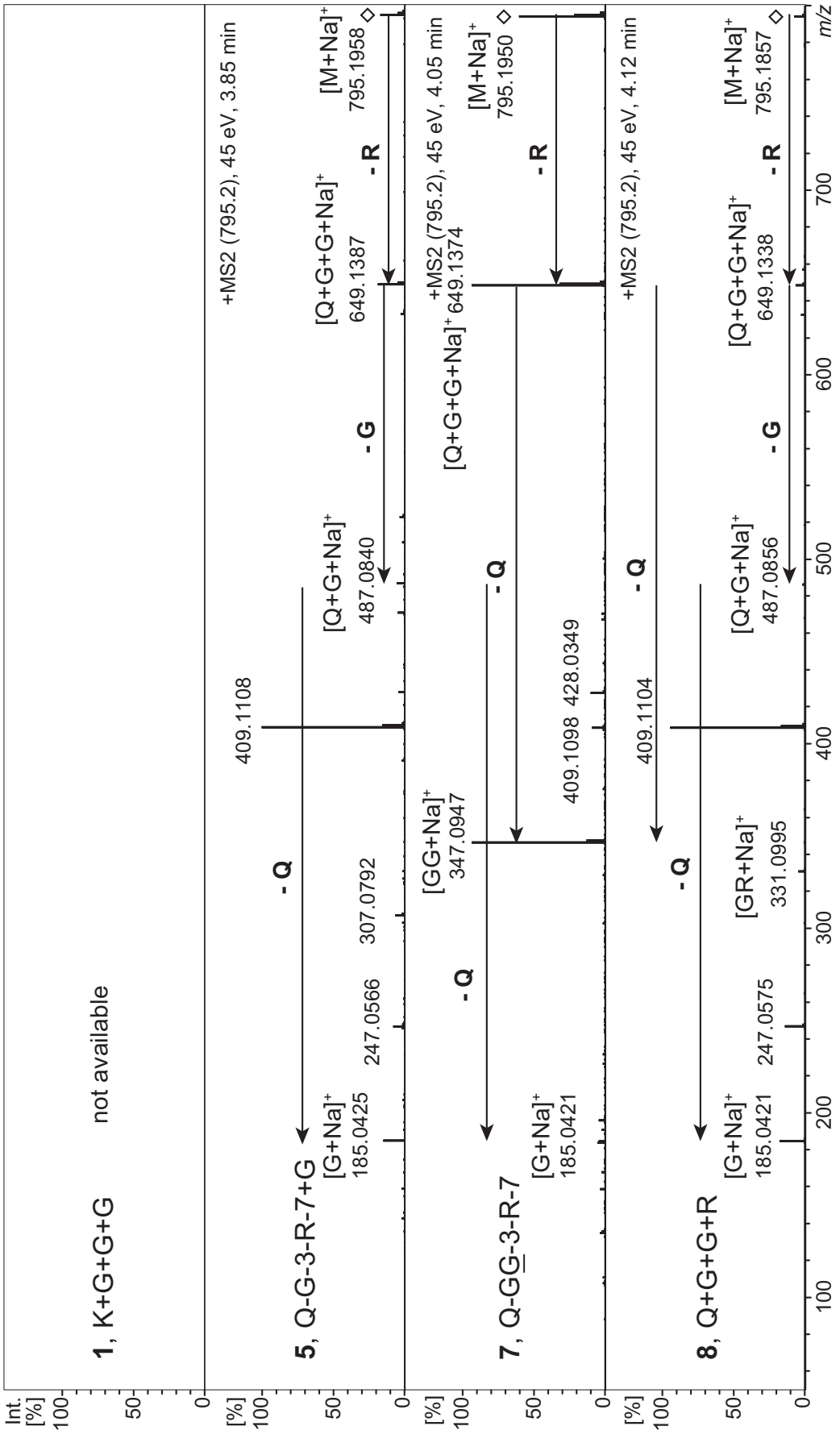
Supplementary figure 53: (+)-ESI-CID-MS/MS of triglycosylated flavonols $C_{33}H_{40}O_{20}$ $[M+Na]^+$ at *m/z* 779: flavonol **2** and **10**.



Supplementary figure 54: (–)-ESI-CID-MS/MS of triglycosylated flavonols $C_{33}H_{40}O_{21}$ $[M–H]^-$ at *m/z* 771: flavonols **1**, **5**, **7** and **8**.



Supplementary figure 55: (+)-ESI-CID-MS/MS of triglycosylated flavonols $C_{33}H_{40}O_{21}$ $[M+H]^+$ at *m/z* 773: flavonols **1**, **5**, **7** and **8**.



Supplementary figure 56: (+)-ESI-CID-MS/MS of triglycosylated flavonols $C_{33}H_{40}O_{21}$ $[M+Na]^+$ at *m/z* 795: flavonols **5**, **7** and **8**.

References

[1] Coutinho, I. D., Baker, J. M., Ward, J. L., Beale, M. H., Creste, S., Cavalheiro, A. J., *J. Agric. Food Chem.,* **2016**, *64*, 21, 4198-4206

[2] Piasecka, A., Sawikowska, A., Krajewski, P., Kachlicki, P., *J. Mass Spectrom.,* **2015**, *50*, 3, 513-532

[3] Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C., Yu, S., Zhang, Q., Luo, J., *Proc Nat Acad Sci,* **2013**, *110*, 50, 20320–20325

[4] Dong, X., Chen, W., Wang, W., Zhang, H., Liu, X., Luo, J., *J. Integr. Plant Biol.,* **2014**, *56*, 9, 876-886

[5] Yang, Z., Nakabayashi, R., Okazaki, Y., Mori, T., Takamatsu, S., Kitanaka, S., Kikuchi, J., Saito, K., *Metabolomics,* **2014**, *10*, 4, 543-555

[6] Matsuda, F., Nakabayashi, R., Yang, Z., Okazaki, Y., Yonemaru, J., Ebana, K., Yano, M., Saito, K., *Plant J.,* **2015**, *81*, 1, 13-23